修士論文

画像内容の記述に基づいた階層的キーワードの生成と利用に関する研究

指導教官　　　　田中 英彦 教授

東京大学大学院 工学系研究科 情報工学専攻
46755　井手 一郎

平成 8 年 1 月 29 日 提出

内容梗概

昨今の驚異的な情報処理環境の向上により、高速度・大容量の情報の送受信および処理が可能になりつつある。そのような状況を背景として、膨大な量の情報を整理して蓄積する VOD (Video On Demand) や電子図書館をはじめとする、マルチメディアデータベースに関する研究が盛んに行われ始めている。

本論文では、画像データベース中での検索の際に用いるデータの検索用キーの適切な構造および、それを用いた検索手法に関して検討する。

そのために、前半では人間による画像の描写の解析を行なう。ここでは、人間が画像中のどのような事物に注目するか、という実験を通じて画像に付与するキーに要求される条件を探る。

後半では、その結果得られた知見をもとに、効率的な検索を実現するために階層的キー「描写木キー」を利用した画像データベースシステムの構築法を提案する。

本研究の特徴は、画像データベースを対象とするために、キーに階層的な構造をもたせたことと、将来の動画像データベースへの適用を念頭におき、名詞に限らず動詞をもキーに含めた点である。

目 次

1 序論 **1**
- 1.1 研究の背景 1
- 1.2 研究の目的 1
- 1.3 本論文の構成 2

2 画像データベース **3**
- 2.1 既存の画像データベースの問題点 3
  - 2.1.1 データベースの生成法に関する問題点 4
  - 2.1.2 データベースの検索手法に関する問題点 4
- 2.2 画像データベースの動向 6
  - 2.2.1 マルチメディアデータベース 6
  - 2.2.2 人間による画像の描写の解析の必要性 7

3 人間による画像の描写の解析 **8**
- 3.1 描写の定義 8
- 3.2 画像と描写 9
  - 3.2.1 描写の階層性 9
  - 3.2.2 描写木による画像記述モデル 9
  - 3.2.3 描写木における描写間の関係の種類 10
- 3.3 人間による画像の描写の特徴 11
  - 3.3.1 描写の多様性と共通性 11
  - 3.3.2 背景知識・注視点と描写 11
- 3.4 描写の一般的特徴に関する実験 13
  - 3.4.1 実験目的 13
  - 3.4.2 実験手順 13
  - 3.4.3 実験に関するデータ 13
  - 3.4.4 実験結果と考察 17
- 3.5 背景知識・注視点と描写との関連に関する実験 30
  - 3.5.1 実験目的 30
  - 3.5.2 実験手順 30
  - 3.5.3 実験に関するデータ 30

3.5.4 実験結果と考察 34
3.6 描写木に要求される条件・特徴 38

4 描写木を利用した画像データベースシステム 39
4.1 システムの構成 39
4.1.1 システムの概略 39
4.1.2 問い合わせ文解析部 40
4.1.3 概念正規化処理部 40
4.1.4 検索処理部 41
4.2 検索処理部における描写木キーの利用法 42
4.2.1 描写木キーの構成 42
4.2.2 描写木キーを利用した画像データベースの検索法 44
4.3 描写木の生成法 46

5 今後の展望 48
5.1 動画像データベースへの拡張 48
5.1.1 動画像の構成 48
5.1.2 描写木の利用法 49
5.1.3 動作の表現 49
5.2 動画像データベースの (半) 自動生成 51
5.2.1 カットの切り替えの検出と事物の抽出 51
5.2.2 動画像データベース (半) 自動生成システム 51

6 結論 53

図 目 次

2.1 マルチメディアデータベース生成の将来像 6
2.2 人間による画像の描写の解析の必要性 7

3.1 描写の階層性 9
3.2 理想的な描写木 9
3.3 is-part-of 関係の例 10
3.4 is-kind-of 関係の例 10
3.5 背景知識・注視点と描写 12
3.6 画像 1 14
3.7 画像 2 15
3.8 描写木 (画像 1) 17
3.9 描写木 (画像 2) 18
3.10 回答数の分布 (画像 1) 20
3.11 回答数の分布 (画像 2) 21
3.12 注視事物の回答数の分布 (画像 1) 22
3.13 注視事物の回答数の分布 (画像 2) 22
3.14 事物の占有領域の面積の割合と回答数との関係 (画像 1) 25
3.15 事物の占有領域の面積の割合と回答数との関係 (画像 2) 25
3.16 離心率と正規化した回答数との関係 (画像 1) 27
3.17 離心率と正規化した回答数との関係 (画像 2) 27
3.18 画像 3 31
3.19 画像 4 32

4.1 描写木をキーとして用いる画像データベースシステム 39
4.2 描写木の構造 –名詞木と動詞リスト– 42
4.3 is-part-of 関係と is-kind-of 関係の違いの例 43
4.4 is-kind-of 関係と is-part-of 関係を区別した概念関係の階層の例 43
4.5 検索の具体例 44
4.6 描写木の自動生成 46

5.1 動画像の構成 (文献 [24] より) 48
5.2 動画像における描写木の変化 49

5.3 動画像データベース (半) 自動生成システム 52

6.1 正規化関数の場合分け 55
6.2 正規化関数の例 (画像 2 の場合) 56

表 目 次

3.1 「全回答」の上位からの回答数の割合と事物数 19
3.2 回答事物数の比較と全回答事物数に対する割合 23
3.3 平均重複率 28
3.4 画像3におけるガ格の回答パターンと回答数 34
3.5 画像3における「狼」の認識 34
3.6 画像4におけるガ格の回答パターンと回答数 36
3.7 画像4におけるニ格の回答パターンと回答数 36

第 1 章

序論

1.1 研究の背景

昨今の情報処理環境の驚異的な向上、とりわけ計算機の処理能力の向上や記憶媒体の大容量化、さらにはそれらの価格性能費の向上による一般家庭への普及、GII (Global Information Infrastructure) 構想をはじめとする情報通信基盤の整備 (文献 [41]) により、21 世紀初頭には国境を超えて大量のマルチメディア情報が氾濫するようになることが予想される。

そのような時代の到来を目前に控え、氾濫するマルチメディア情報の中から利用者が必要とするものを選別して提示する必要が高まりつつある。そのため、膨大な量のマルチメディア情報を検索しやすいように整理して蓄積する、VOD (Video On Demand) や電子図書館 (文献 [39], [40]) をはじめとする、マルチメディアデータベースに関する研究や開発が盛んに行われ始めている。

その中には、マルチメディアデータに適した伝送方式やデータを蓄積するためのハードウェアに関する研究も含まれているが、より根本的な問題である、効率的な検索を実現するためのキーが備えるべき条件や特徴に関する研究はあまり行われていない。

1.2 研究の目的

本論文では、画像データベースの効率的な検索を実現するための基礎的研究として、おもにキーに注目した議論を行う。

前半では、人間による画像の描写の解析を通じて、画像データベースの検索の際に重要な役割を果たすキーの、画像を扱うのに適切な構造を検討する。

後半では、その結果得られた知見をもとに、効率的な検索を実現するために導入する、画像内容の記述に基づいた階層的キー「描写木キー」を用いた画像データベースを提案する。

これらの議論により、将来のマルチメディアデータベースの効率的検索の実現への足掛かりとすることが本研究の目的である。

1.3 本論文の構成

まず、第2章では、既存の様々な方式の画像データベースの問題点を簡単にまとめ、マルチメディアデータベースの中における本研究の位置づけについて述べる。

第3章では、人間による画像の描写を解析し、階層構造のキーワード「描写木」を念頭に置き、その特徴や要求される条件について分析する。

第4章では前章で提案した描写木を用いた画像データベースシステムの構築法および利用法を検討する。

第5章では以上の議論を踏まえて動画像に対応させるための今後の展望を述べ、最後に第6章で結論をまとめる。

第 2 章

画像データベース

2.1 既存の画像データベースの問題点

1.1 で述べたような現状を背景として、画像データベースに関する研究が活発に行われている。実際、VOD などは実用的なサービスのためのシステムも開発されており、米国の一部地域などでは商用サービスも行われ始めている。

しかし、現在の画像データベースでは、内容に関する情報として、たとえば VOD の場合、あらすじ, 題名, 出演者などの簡単なキーワードという、非常に大雑把で限定的なものしか用意されていないのが普通である。これらの情報は画像の内容に深く踏み込んだものではなく、シーン単位での検索を行うような本格的な番組の選別を行うためには不十分であり、近い将来要求されるようになると思われる水準を質的に満たすことができない。また、付与されている情報にしても、大半は人手で付与しているのが現状であり、増え続ける需要に量的に応えることも困難である。

もっとも、研究段階のシステムには、検索者が類似画像や概図を入力してそれらに近い画像を検索するようなもの (文献 [10], [12]) や、形容詞や副詞をキーワードとして入力し、画像の色彩的特徴と比較して検索を行うようなもの (文献 [12]) がある。このような、画像による検索や感性的なキーワードによる検索では、画像から定量的な特徴量を比較的容易にかつ自動的に抽出することが可能であり、前述の自動化の問題はおおむね解決される。

しかし、これらのシステムは特定の種類の画像の検索に限られており、一般には画像中の事物に関してあらかじめ形状や色彩について知らずに、言葉 (おもに名詞) によって検索する需要が多いため、その手法を一般的な画像データベースシステムに適用することはできない。

以下の二節では、画像データベースの生成法および検索法に関する問題点を、既存の研究を紹介しながら、さらに検討する。

2.1.1 データベースの生成法に関する問題点

すでに実用化されているものを含め、既存のほとんどの画像データベースでは、画像データに検索用のキーワードが付与されている。しかし、対象を限定しない一般的な画像認識は非常に困難なため、キーワードの付与は通常人間が画像を見て、重要と思われる事物や主題を抽出したり、映画やテレビドラマなどであれば、出演している俳優の名前を抽出したりして行っている。

これらのキーワードは必ずしも画像の内容に深く踏み込んだものではないため、きめ細かい検索が実現できない可能性があるという質的な問題がある上に、今後 **1.1** で述べたような時代背景により、提供される情報量が増えるにつれ、人手に頼ることは量的にも困難となることが予想される。

そのため近年、画像のライブラリ化の手法として、画像単体ではなく、付随する言語情報を利用して不足している情報を補うようなものが注目されている。既存の多くの画像認識システムでは、システム側に認識する事物に関するかなり詳細なモデル (知識) を事前に用意しておかなければならなかった。それに対し言語情報を利用する手法では、画像に付随する言語情報を利用することにより、システム側では最低限の画像モデルを用意すれば済むため、画像処理による認識が大幅に軽減される。もっとも、データとなる画像に言語情報が付随していなければこのような手法は適用できないが、実際にテレビジョン放送などで提供される画像には音声や字幕などの言語情報が付随することが多く (文献 [23], [26], [33]) 、雑誌, 新聞や図鑑などの中の写真には解説文が付随することが多い (文献 [1], [9], [18], [20], [21])。

本研究では、近い将来需要が拡大すると思われる画像データに付与するキーワードの (半) 自動生成を行うための基礎的研究として、キーワードに要求される条件を洗い出し、画像から適切なキーワードを (半) 自動生成する技術を確立するための足掛かりとする。

2.1.2 データベースの検索手法に関する問題点

検索手法とは、検索者がデータベースに対して問い合わせをする手段のことである。現在採用されたり提案されたりしている手法は、大きく分けて次の二種類に分類できる。

1. 自然言語による検索
   検索者はキーボードや音声入力装置などを用いて自然言語 (単語や文章) により検索を行う。
   この手法は、検索キーの種類によりさらに次の二種類に分類できる。
   (a) データに自然言語のキーが付与されているもの
      問い合わせの自然言語 (文, 単語) とキーとの照合を行う。
      [長所] 検索の際に画像処理を行う必要がないため、処理が非常に簡単である。

  [短所] おもに問い合わせの多様性・多義性により、単純な照合では効果的な検索は行えないので、概念辞書などを用いた多様性の解消が必要である。
  また、現時点ではキーの付与は人手に頼るところが大きい。

   (b) データに自然言語以外のキー (データの何らかの特徴量) が付与されているもの
   検索文字列に何らかの変換を施し、キーとの照合を行う。
   色彩情報などを扱った感性的な検索を行うシステムに多い。

  [長所] 形容詞や副詞による感性的な検索が可能である。
  また、データから直接特徴量を抽出するため、人手による処理が不要である。

  [短所] 事物 (名詞など) による検索ができない上に、形容詞や副詞による感性的な検索においては、語の定量化が課題である。

2. 画像による検索
検索者はマウスやイメージスキャナなどの入力装置を用いて、画像を入力して検索を行う。
この手法は、画像の扱い方により次の二種類に分類できる。

   (a) 画像中の事物を直接扱うもの
   問い合わせとして入力された画像 (実画像や手書きの概図など) との照合を行う。

  [長所] 画像中の事物を画像として直接検索できる。
  また、キーの付与が不要なため、人手による処理が不要である。

  [短所] 単純な照合を行うのでさえ画像処理が困難である。
  実用的なシステムにする際に、場合によっては高度な画像認識を行う必要があり、詳細な画像モデルを用意しなければならず、非常に限定された対象しか扱えないシステムになる。

   (b) 画像中の事物を直接扱わないもの
   問い合わせとして入力された画像とデータの何らかの特徴量との照合を行う。

  [長所] 画像の全体的な雰囲気 (「穏やかな」など) や色調 (「赤っぽい」など) による感性的な検索が可能である。
  また、キーの付与が不要なため、人手による処理が不要である。

  [短所] 画像中の事物は扱わないので、具体的な事物 (の画像) による検索ができない。

このようにして見ると、最も需要があると思われる、具体的な事物による検索を行うためには、1.(a) や 2.(a) の手法を用いるのがよいと思われるが、2.(a) の短所である画像認識の困難さは非常に強い制約である。

そこで本研究では、画像データベースにおいても有効と思われる 1.(a) の手法を扱う。従って以下では、1.(a) の短所を克服するために、キーの付与の (半) 自動化および多義性を解消した検索法に関して検討する。

2.2 画像データベースの動向

2.2.1 マルチメディアデータベース

マルチメディアデータベースとは本来複数の媒体、すなわち画像・音声・文字列などを融合したデータベースのことを指すが、現時点では主に画像と文字列を扱い、若干の音声が付随する程度の画像データベースシステムを指すことが多いようである。しかし、将来的には図 2.1に示すように、ケーブルテレビジョン (CATV) などが時々刻々と提供する大量のマルチメディア情報を処理し、(半) 自動的にデータベースに蓄積するようなシステムが要望され、また登場することになると予想される。

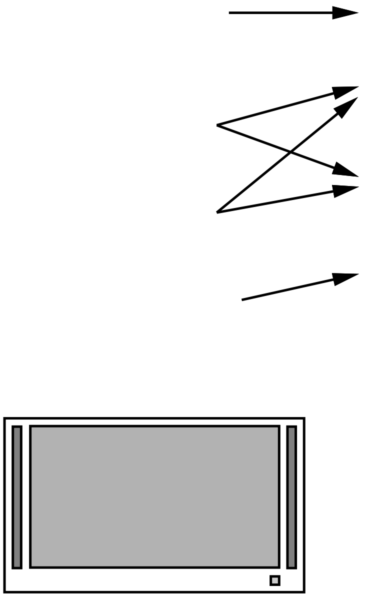

図 2.1: マルチメディアデータベース生成の将来像

このようなシステムを構築する際には図 2.1のシステム中に示すような様々な要素技術が必要となる。現時点ではこれらの技術の大半は計算量, 汎用性や精度などの点で実用に耐えない。ことに、不特定話者による会話の音声認識, 一般的な画像認識や会話文の解析は非常に困難である。

しかし、本研究ではこれらの要素技術の発展は待つこととして、既存の技術でもある程度実現可能と思われる、自然言語処理と画像処理が互いに欠けている情報を補いあうような画像データベースシステム[1]の生成を目標とする。

---

[1] 音声はすでに文字列に変換されたものとして扱う。

2.2.2 人間による画像の描写の解析の必要性

本研究では、図 2.2に示すような仮定に基づき、データに付与する適切なキーに関する知見は、人間による画像の描写の解析を通じて得られるものとする。

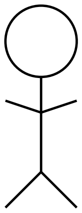

図 2.2: 人間による画像の描写の解析の必要性

図 2.2に示した内容を簡単に説明する。

1. 検索の際には、検索したい画像のイメージ$_{\mathbf{(A)}}$が検索者の頭の中に (漠然と) 存在する。
2. 検索者は **(A)** を言葉で表現（ = 描写 ）し$_{\mathbf{(B)}}$、データベースへの問い合わせとする。
3. データベースシステムは、問い合わせの条件を満たすキー$_{\mathbf{(D)}}$をもつデータを検索する。
4. 検索結果$_{\mathbf{(C)}}$は、**(A)** に近いほどよい。

→ データの **(D)** が **(B)** に近いほどよい。

  → 良い **(D)** を生成することが、良い画像データベースの構築につながる。

  → 良い **(D)** に要求される条件は、**(B)** を解析することにより知見が得られる。

∴ 人間による画像の描写の解析が必要である。

このような動機のもとに、次章では二つの心理実験を通じて、人間による画像の描写の解析を行う。

第 3 章

人間による画像の描写の解析

前章では、人間による画像の描写を解析する必要性に関して述べたが、本章では、前半で人間による画像の描写に関するモデルを提案し、後半では実際に行った心理実験とその解析を通じて、前半で提案したモデルの正当性を確認する。

3.1 描写の定義

すでに、「描写」という言葉を未定義で使用してきたが、以降の混乱を防ぐため、ここで改めて定義をしておく。

基本的には、「犬が道を歩いている」というように、人間が画像を見てそこに映っている事象を説明するために自然言語で表現することを「描写」と呼ぶ。

しかし、実際に計算機で処理する際には前に示した例文のような自然言語文のままではデータとして扱いにくいので、下に例示するように、文中の名詞や動詞を抽出して関連付けを行なったデータ構造のことも「描写」と呼ぶこととする。このような関係付けは、形態素解析システムや格解析システム[1]を用いることにより、会話文などの文法規則から逸脱したような文や係り受け関係が複雑な文でなければ、比較的容易に実現可能である。

$(ex.)$ 「犬が道を歩いている」 → 歩く (犬, 道)

なお、本章ではおもに、描写を代表してガ格 (主格) をとる名詞 (上に示した例では「犬」) を中心として議論する。

---

[1] 実際に計算機による (半) 自動処理システムを開発する際には、京都大学工学部の長尾研究室および奈良先端科学技術大学院大学の松本研究室で開発された JUMAN と呼ばれるシステムを用いることを考えている。

3.2 画像と描写

本節では、描写の特性を踏まえた上で、画像を記述するために描写木という階層構造のモデルを提案し、その特徴について述べる。

3.2.1 描写の階層性

一般的に考えて、図 3.1に示すように、ある描写はより細かい描写に分解できるので、全体として階層構造をとると考えられる。

図 3.1: 描写の階層性

3.2.2 描写木による画像記述モデル

どのような画像も、理想的には描写の集合によって記述し尽くされると仮定する[2]。言い換えれば、描写の集合による記述を画像と等価なものと仮定する。

さらに、前項で述べた階層性を併せて考え、画像は描写の階層構造により等価に表現することができると仮定し、このような構造を「(理想的な)[3]描写木」と呼ぶ。

図 3.2: 理想的な描写木

---

[2] 現実にはそのようなことは不可能だし、たとえ可能であっても可算個の描写で尽くされる保証はない。

[3] 後出の描写木はかならずしも画像と「理想的に」等価なものではない。

3.2.3 描写木における描写間の関係の種類

描写相互間の関係は以下の三種類に分類できる。ただし、必ずしも一意に特定の関係に決定できるとは限らない。

is-part-of 関係

図 3.3に例示するように、事物の物理的な包含関係のことである。

図 3.3: is-part-of 関係の例

is-kind-of 関係

図 3.4に例示するように、事物の概念的な包含関係のことである。

図 3.4: is-kind-of 関係の例

is-a 関係

is-part-of 関係や is-kind-of 関係とは少々異なり、物理的・概念的に同一階層にある描写間の関係である。

本研究で扱う階層関係では基本的に、各描写はすでに同一階層にある is-a 関係を吸収しているものとして扱う。

3.3 人間による画像の描写の特徴

次節では人間による静止画像の描写に関する心理実験を行なうが、本節ではその前に人間による画像の描写がどのように行なわれるかを、描写木を用いたモデルを通して検討する。

3.3.1 描写の多様性と共通性

人間による画像の描写には、多様性と共通性が存在する。多様性や共通性が生じる理由としては、次の二つの要素が大きく関係すると思われる。

- 背景知識
  個人個人が現時点までに獲得した知識の集合。
- 注視点
  画像を見る際に、興味をもつ対象 (事物) 。

次項ではこれらが描写にどのようにして影響を及ぼすかを、簡単なモデルを用いて検討する。

3.3.2 背景知識・注視点と描写

ここでは、画像と等価な理想的な描写木を用いたモデルにより、背景知識・注視点によって人間による描写が変わる様子を示す。
図 3.5に背景知識と注視点、さらにそれらによって描写がどのように形成されるかを表すモデルを示す。

図 3.5のモデルを簡単に説明する。

- 背景知識
  人間のもつ背景知識とは、現時点までに獲得されてきたあらゆる情報のことであり、個人個人で異なる。
  しかし、兄弟姉妹、同程度の教育水準の人、さらには同民族というように、文化的・社会的に共通の基盤をもつ人の間では背景知識にも何らかの共通性が見られると推測される。
- 注視点
  人間はもっている背景知識を常に総動員しているわけではなく、内的要因 (心理状態など) や外的要因 (周囲の状況など) により形成されるその時々の興味によって注目する対象 (注視点) が変わると思われる。
  このように、注視点は各人で異なるだけでなく、その時々の状況によっても異なる。
  しかし、個人差の少ない注目対象 (色彩や形状が特異なもの、内容的に衝撃的な事象など) もあり、視覚的・内容的な要因による共通性もあると思われる。

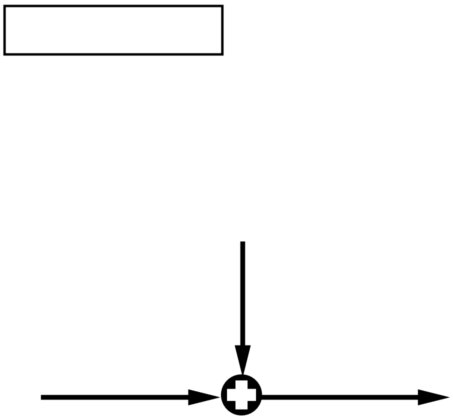

図 3.5: 背景知識・注視点と描写

以下の二節では、本節で取り上げたモデルのもとに、心理実験を行ない、人間による画像の描写の特徴を明らかにする。

3.4 描写の一般的特徴に関する実験

3.4.1 実験目的

以下の事項を目的として実験を行なった。

1. 人間が画像中のどのような事物に特に注視するかを調べる
2. 人間が画像中の事物にどこまで注視するかを調べる
3. 人間による描写の多様性と共通性を調べる
4. 得られた結果から実際に描写木を作成する

3.4.2 実験手順

二枚の白黒の静止画像に関して、以下の手順で実験を行なった。
用いた画像を図 3.6と図 3.7に掲載する。
また、実験は個人面接方式で行ない、随時必要に応じて回答に関する質問をした。

1. 被験者に画像を **15** 秒間見せる
2. 画像を隠し、印象に残っている事物を三つ挙げさせる
   実験目的 1. に関するデータを得るため
3. 再び画像を見せ、**2.** で挙げた三つの事物に関して、一行程度で簡単に描写させる
   実験目的 3. に関するデータを得るためと、動詞に関するデータを得るため
4. 画像中の描写に値すると思われる事物を思いつく限り記させる[4]
   実験目的 2. と 4. に関するデータを得るため

3.4.3 実験に関するデータ

被験者

被験者に関するデータは以下の通りである。

- 人数 ... 男性 20 人 + 女性 5 人 = 合計 25 人
- 年齢 ... 20 歳代
- 職業・専門分野
  大半は東京大学の工学系 (専門分野は電気工学, 通信工学, 情報工学など) の大学生や大学院生

---

[4] ここで挙げさせた事物には実験手順 1. で挙げさせた三つを含む

画像 1

©1995 アサヒグラフ

図 3.6: 画像 1

画像 2

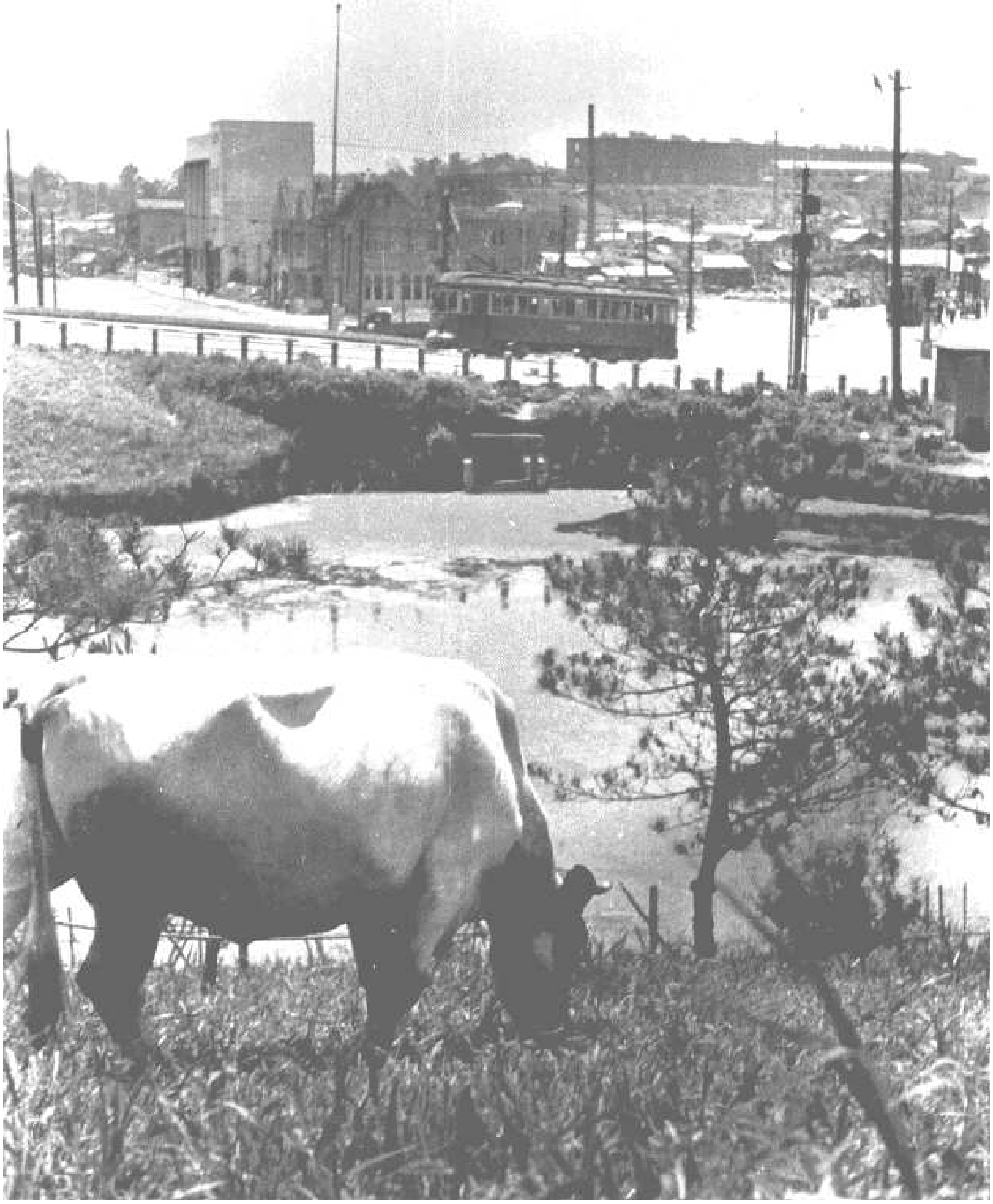

図 3.7: 画像 2

回答用紙

実際に実験に用いた回答用紙を示す。

描写実験記入用紙

氏名: ____________________

実験日: 平成 7 年　　月　　日

描写画像: □番

1　印象に残った事物 (=名詞) を 3 つ挙げて下さい

1. ____________
2. ____________
3. ____________

2　1 で挙げた事物について一行程度で簡単に描写して下さい

注: 「○○ (=1 で挙げた事物) が (△△を) □□ている」という文型にして下さい。

1. ____
2. ____
3. ____

3　描写に値すると思われる事物を好きなだけ挙げて下さい

____________ ____________ ____________ ____________

____________ ____________ ____________ ____________

____________ ____________ ____________ ____________

____________ ____________ ____________ ____________

____________ ____________ ____________ ____________

3.4.4 実験結果と考察

描写木の例

回答をもとに作成した画像 1, 画像 2 の描写木を各々、図 3.8, 図 3.9に示す。

なお、ここに示した描写木は、実験の結果をもとにして、筆者が階層の上下関係や包含関係 (is-kind-of 関係と is-part-of 関係の区別など) を判断して構築したものであり、唯一無二の構造というわけではない。

また、is-a 関係にあるものや漢字の表記の違いなどは各事物に吸収してある。

図 3.8: 描写木 (画像 1)

図 3.9: 描写木 (画像 2)

集計結果

実験手順 2. の集計結果である「注視事物」の度数分布と、実験手順 4. の集計結果である「全回答事物」の度数分布を示す。

実験手順 4. の回答は、被験者に必要なだけ長い時間画像を見せた結果であるので、その被験者が認識した画像中のほぼすべての事物が列挙されているものと考えられる。さらに、全被験者の回答を集計した「全回答事物」は、人間が一般に認識し得るほぼすべての事物の集合であると考えられる。

一方、「注視事物」は、画像を 15 秒という比較的短時間しか見せず、しかも記入の際に画像を隠しているので、強く印象に残った事物が列挙されているものと考えられる。

比較的短時間と言っても、15 秒という視覚系の処理を経て脳が高度な理解を行なうのに十分な時間では、単に視覚的に目立つ事物のみが印象に残るのではなく、内容的に目立つ事物も印象に残る。本実験では、人間の視覚的特性を求めるのが主題ではなく、むしろ背景知識などとの関連を示すことを目指すので、このような、脳による理解に必要とされる時間よりも長い時間に設定した。

回答数の分布

図 3.10と図 3.11に画像 1, 画像 2 の回答数の分布を示す。これらは、回答された事物から is-a 関係および is-kind-of 関係にある事物を代表的な事物に吸収した結果残った事物の回答数の分布である。

前に述べたように、全回答事物は画像中のほぼすべての事物が挙げられていると考えられるが、その中でも回答数の分布で上位にある事物は画像のキーワードとして不可欠なものである。表 3.1に、全回答事物ののべ回答数の上位から 10%刻みの累積回答数を挙げる。

割合	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
画像 1	1	2	3	5	7	9	12	17	23	40
画像 2	0	1	2	4	5	7	9	13	18	33

表 3.1: 「全回答」の上位からの回答数の割合と事物数

このようにして見ると、のべ回答数の上位から 80%の回答数を占める事物を含むためには、画像 1 では$\frac{17}{40} = 42.5\%$、画像 2 では$\frac{13}{33} \simeq 39.4\%$と、約四割の事物で十分であり、意外と少ないことが分かる。これは、回答数の分布が下位方向に長く尾を引いた形状をしていることからも直観的に分かる結果である。

これらのグラフの特徴として、「全回答事物」には回答数が少ない事物が多いことが挙げられ、被験者個人個人による多様性の大きさを示している。

その一方で、いくつかの回答数が多い事物に被験者のほとんどが回答をしているのを見ると、同時に共通性も存在することが分かる。

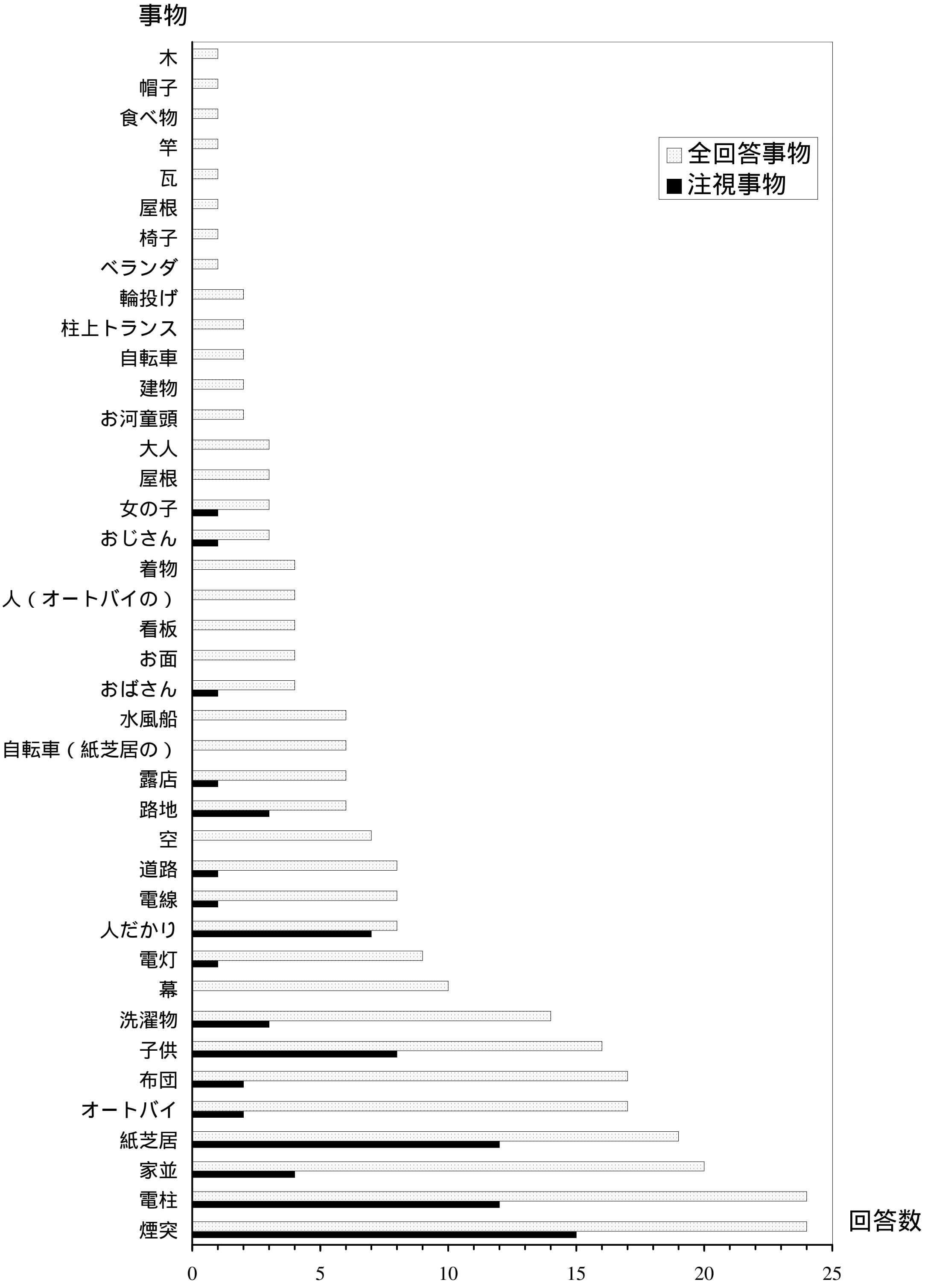

図 3.10: 回答数の分布 (画像 1)

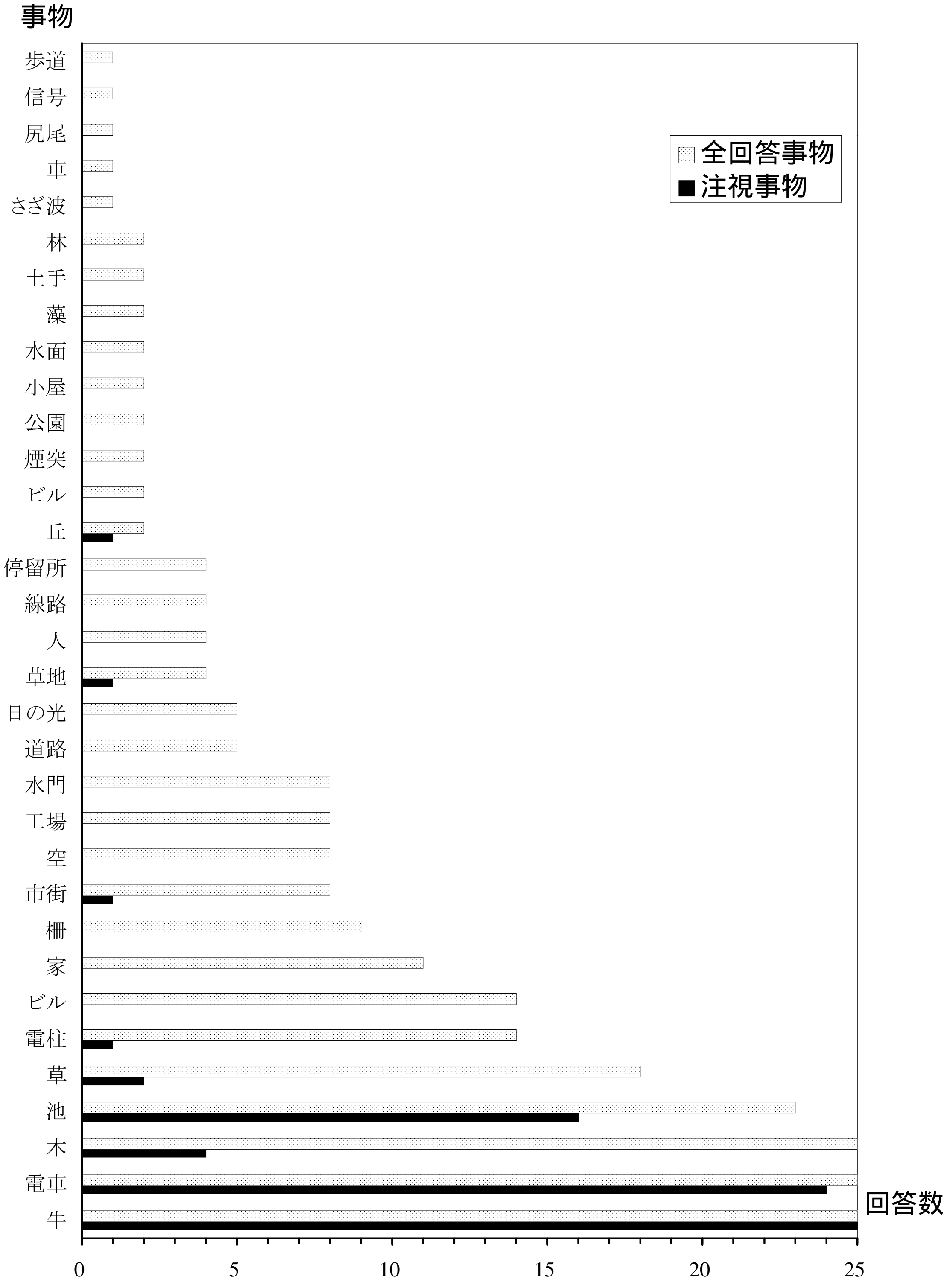

図 3.11: 回答数の分布 (画像 2)

ここで注目されるのは、「全回答事物」と「注視事物」とでは回答数の順位が必ずしも一致しないことである。これは、前者では回答数と回答時間に制限を設けなかったことにより、さほど目立たない事物や、内容的にさして目立たなくとも視覚的に目立つ事物が挙げられたためと思われる。

注視事物

図 3.10や図 3.11では全回答事物の度数の多い順に整列したグラフを示したが、ここでは注視事物に注目するために、注視事物のみの回答数の分布を図 3.12と図 3.13に示す。

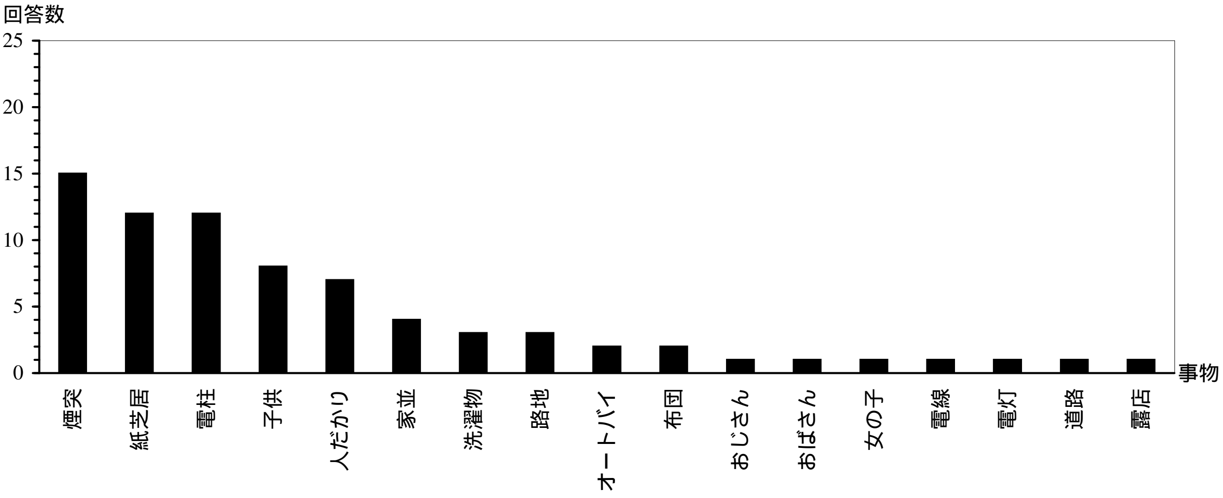

図 3.12: 注視事物の回答数の分布 (画像 1)

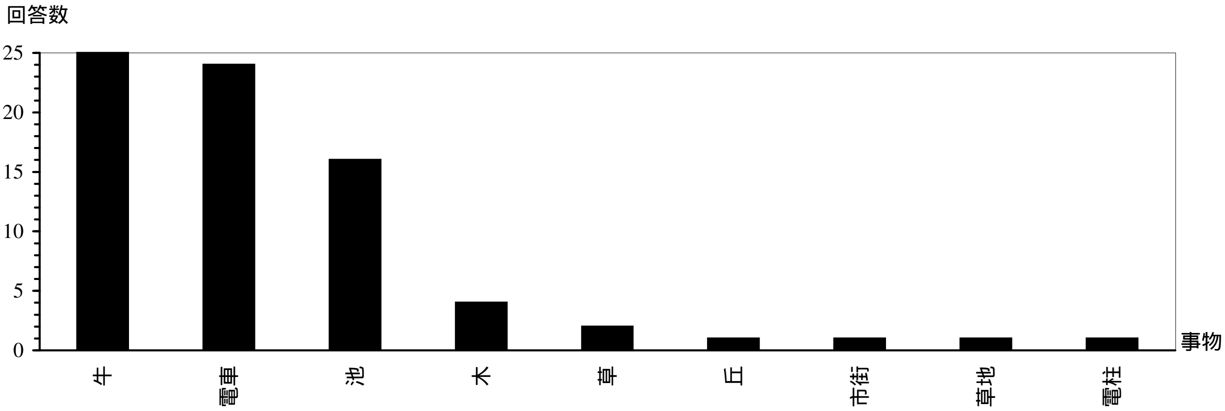

図 3.13: 注視事物の回答数の分布 (画像 2)

「全回答事物」が画像中のあらゆる事物を列挙したものであるのに対し、「注視事物」は、視覚的・内容的に目立つ事物であり、その画像のキーワードとして適切なものとなる。

注視事物数と全回答事物数を比較するために、表 3.2に、回答された事物数の比較を示す。

画像	全回答事物数	注視事物数	主注視事物数[5]
画像 1	40	17 (43%)	10 (25%)
画像 2	33	9 (27%)	5 (15%)

表 3.2: 回答事物数の比較と全回答事物数に対する割合

このように、全回答事物数に占める注視事物数の割合は、両画像で半分以下に収まるとは言えるものの、実際の数値としてはかなり異なる。これは分母となる全回答事物数、つまり画像の複雑さや、分子となる注視事物数、つまり視覚的・内容的に目立つ事物の多さが反映されていると思われる。実際に、画像 1 は画像 2 と比較して全体的に多くの事物が映っている複雑な画像であり、逆に画像 2 は画像 1 と比較して視覚的に目立つ事物が多いと言えそうである。

しかし、それでも注視事物数は全回答事物数の 50%以下、多様性を若干除いた主注視事物数は全回答事物数の 25%以下に収まる。注視事物は人間が見たときに視覚的・内容的に目立つ事物なので、これらを画像に対するキーワードとすれば、画像データベースを検索した際に提示される画像は検索者を満足させるようなものとなり得るだろう。

回答数と画像的特徴との関係

さて、本研究では将来的には画像データベースの生成、つまり画像データに対するキーワードの (半) 自動的な付与を念頭においている。そのためには、実験結果の注視事物の上位にあった事物のように、画像中から視覚的・内容的に目立つ事物を抽出する必要がある。内容的に目立つ事物を画像処理のみによって抽出することはできないが、視覚的に目立つ事物ならば、何らかの画像的特徴量を抽出して (半) 自動生成に寄与できるものと考えられる。

そこで、ここでは画像的特徴量と回答数との関係を調べる。

一口に画像的特徴量と言っても、次に列挙するように様々なものが考えられる[6]。

- 色彩的特徴量
  - 領域の代表色
  - 領域と背景との色の違い

---

[5] 主注視事物とは、度数が一のものを除いた注視事物を指す。

[6] 文献 [36] で佐藤らが用いている特徴量を一部参考にした (文献 [36] の本文には未収録) 。

- 幾何的特徴量
  - 領域の面積
  - 領域の位置
  - 領域の外接長方形の形状
  - 領域の外周長
  - 領域の円らしさ
- その他
  - 領域のテクスチャ

ここでは、これらのうち、本実験結果と関わりの強そうな特徴量である、事物の占有領域の面積と位置を測定し、回答数との比較を行なった。

回答数と面積との関係

図 3.14と図 3.15に、実験において回答された事物の度数とその事物の占有領域が画像全体に対して占める面積の割合をプロットしたものを示す。

占有領域の割合が大きな事物が注目されやすいと予測していたが、これらのグラフを見る限り、必ずしもそのような傾向は見られない。むしろ、図 3.14に顕著なように、占有領域の面積が小さい方がかえって注目されやすい傾向さえ見られる。

それでも、画像 2 における「牛」のように、明らかに大きいために注目されやすかったと思われる事物もある。また、視覚的にあまりに目立たない事物に注目することはない。

改めて考えると、背景や空などは占有領域の面積の割合は大きいが内容的には目立たず、小さくても内容的に目立つ事物の方が優先的に描写されるということが言える。

これらの結果より、人間は単に視覚的に目立つ事物よりも、視覚的にある程度目立ち、かつ内容的に目立つ事物に注視する傾向が強いことが言える。

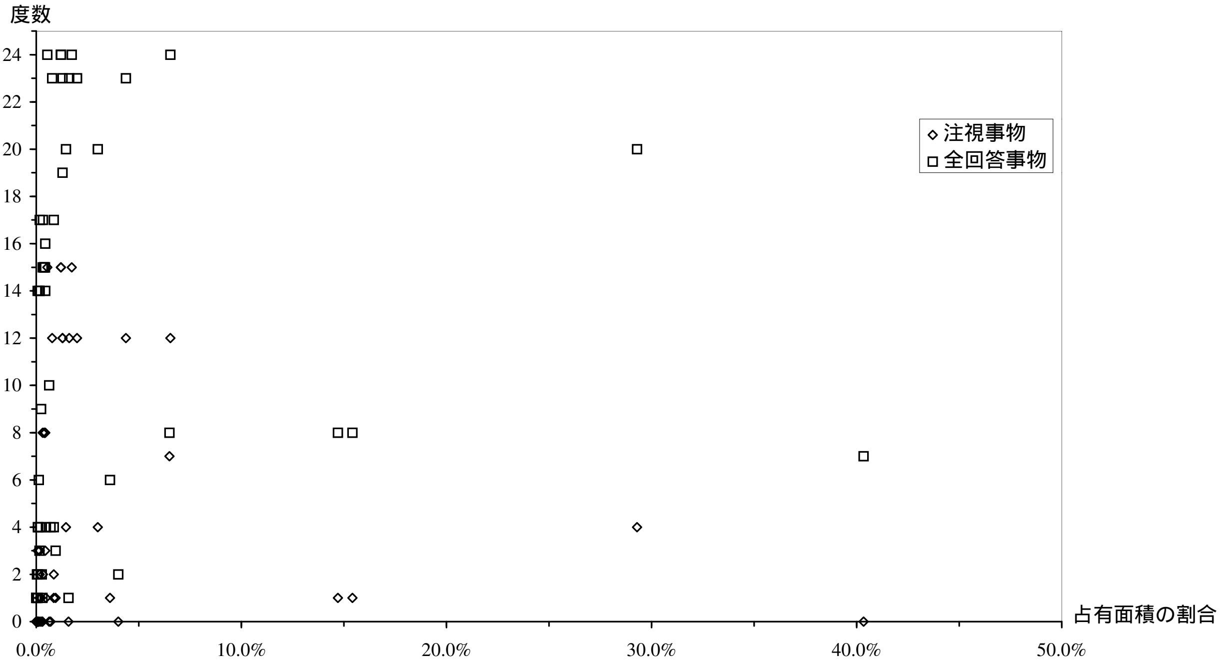

図 3.14: 事物の占有領域の面積の割合と回答数との関係 (画像 1)

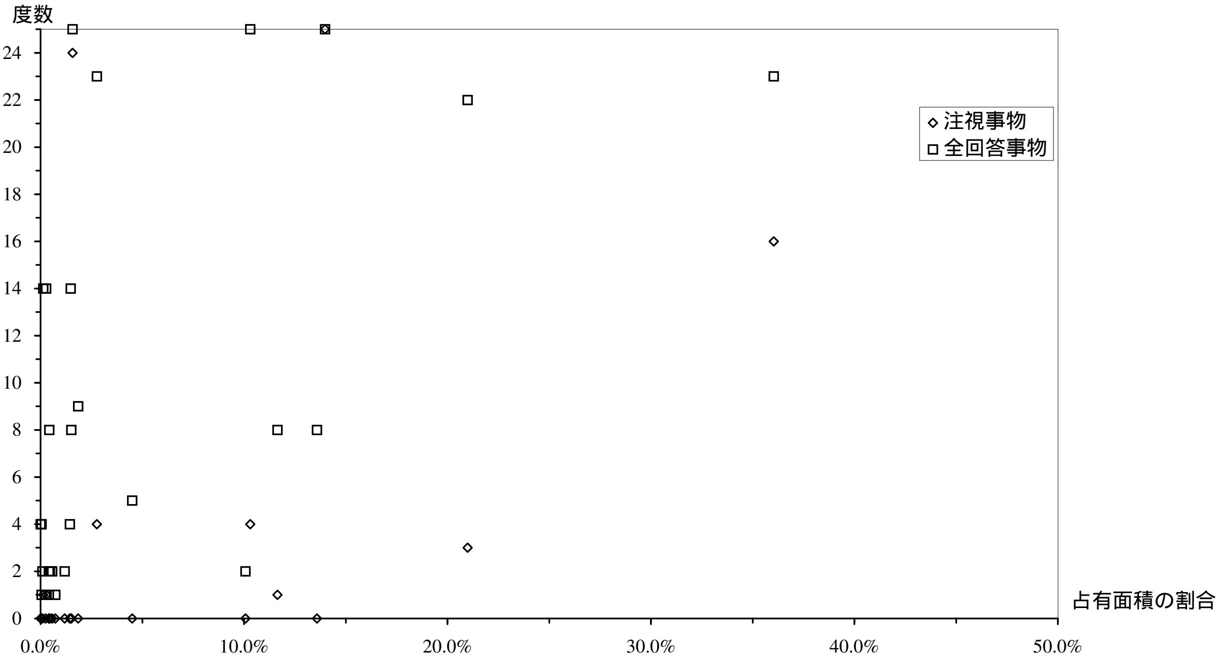

図 3.15: 事物の占有領域の面積の割合と回答数との関係 (画像 2)

回答数と位置との関係

図 3.16と図 3.17に、実験において回答された事物の度数とその事物が画像の中心からどれだけ離れているか (離心率と呼ぶ) [7]との関係をプロットしたものを示す。

なお、離心率 $e(x,y)$ は以下のように定義する。ただし、(x,y) を事物の占有する領域の重心の座標、(X,Y) を長方形の画像の横と縦の長さとする。

$$e(x,y) = \sqrt{\frac{(2x-X)^2 + (2y-Y)^2}{X^2+Y^2}}$$

縦軸には正規化した度数をとった。これは、事物が画像中に一様に散らばっていると仮定すると、画像の中心を中心とするドーナツ状の微小領域について考えるとき、半径に比例して事物が微小領域上に存在する確率が高くなることから、その確率に応じて度数を正規化する必要があるためである。ただし、画像の形状は長方形なので、実際には若干複雑な関数[8]を用いることになる。

さて、図 3.16と図 3.17を見ると、全回答事物は、グラフの理論的な上限[9]である「正規化最大値」の曲線にほぼ沿って分布している。全回答事物は被験者に必要なだけ長い時間画像を見せて回答させたため、さほど目立たない事物を含め、画像全体に満遍なく回答が分布した結果だろう。

一方、注視事物の方は、中心に近い領域では、度数こそ少ないが、全回答事物とそれほど差のない分布をしている。しかし、周縁部 (離心率が 100%に近い方) では注視事物は「正規化最大値」に沿わずに減少傾向にある。これは、周縁部に行くにつれ事物が注視されにくくなることを示唆している。正規化処理を行なっているため、周縁部の領域の狭さによるものではないので、中心に近い領域に存在する事物が注視されやすい傾向があると結論付けられる。

---

[7] 本来ならば事物の占有領域に関する面積分をする必要があるが、ここでは便宜上領域の重心をその領域の代表点として計算した。

[8] 正規化関数に関しては、本筋とは関係ないので、付録に別に記す。

[9] 正規化関数に全被験者数 (25 人) を代入したもの。

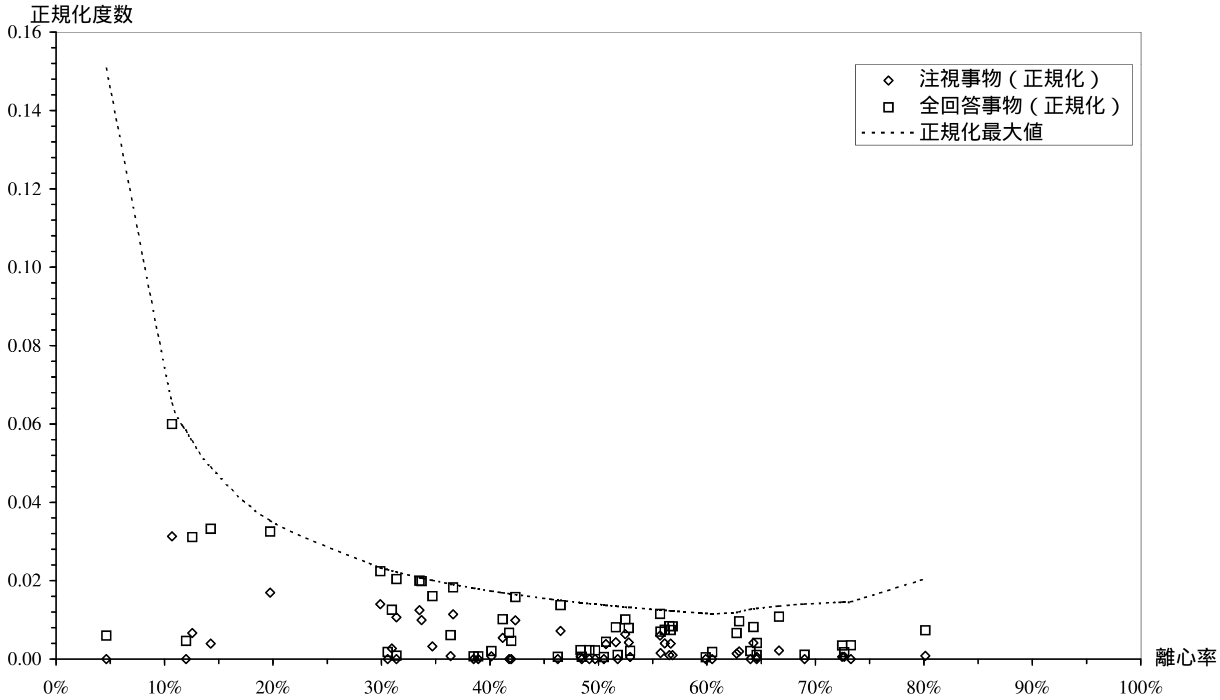

図 3.16: 離心率と正規化した回答数との関係 (画像 1)

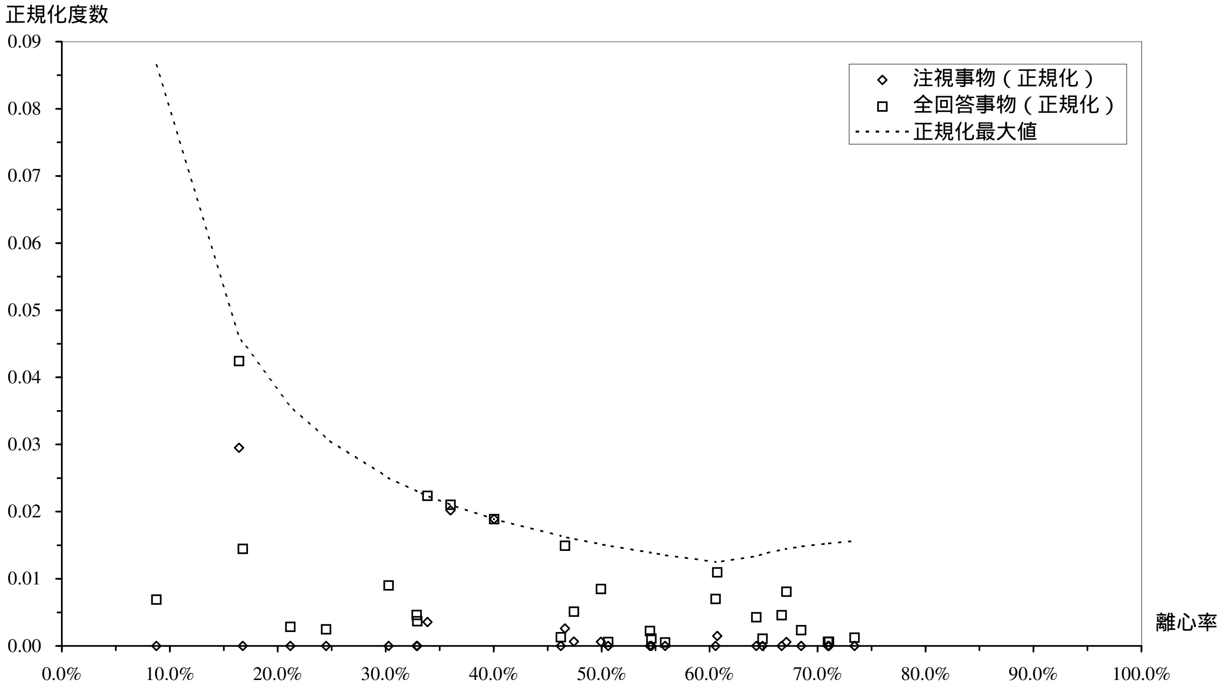

図 3.17: 離心率と正規化した回答数との関係 (画像 2)

人間が描写する事物

図 3.10, 図 3.11, 図 3.12, 図 3.13を見ると、回答は特定の事物に集中していた。集中する理由としては、前項で解析したような画像的特徴のほかに、内容的特徴もあると思われるが、これに関しては画像の解析からは直接知見が得られず、人間による画像の認識に関してより詳細に検討する必要がある。

一方、特に全回答の方では、回答数が一ないし二という事物もかなり存在した。

このように人間による描写には共通性と多様性があり、以下で簡単に検討する。

まず、回答の全体的な共通性を示す指標として、延べ回答数を回答事物数で割った平均重複率を表 3.3に示す。

画像	全回答	注視点
画像 1	6.9 (275/40)	4.4 (75/17)
画像 2	7.4 (245/33)	8.3 (75/ 9)

表 3.3: 平均重複率

表 3.3を見ると、回答の共通性の強さが分かる。平均重複率は画像 1 では若干低いが、画像 2 ではかなり高くなっている。これは、画像 1 は複雑で込み入った画像であるのに対して、画像 2 は比較的簡単な画像であることによると思われる。

次に、回答の多様性に関して見てみる。

次節で行なう実験では個人の背景知識と描写との関連を扱うが、実際に個人の背景知識を直接抽出することは不可能に近い。しかし、共通の社会環境や教育水準、職業や専門などにより形成される局所的な背景知識の共通性により被験者を分類し、比較することは可能である[10]。

以下に、実験結果の中で被験者の背景知識が描写に影響を与えていると思われる興味深い回答例をいくつか示す。

- 画像 **1** で「柱上トランス (配電盤) 」と回答
  電気工学 (高電圧) を専攻する被験者二名のみが回答
- 画像 **1** で「お河童頭」と回答
  女性の被験者二名のみが回答
- 画像 **1** で「紙芝居」に注目はするが、認識できない被験者が **20%**以上存在
  紙芝居を見たことがない、あるいはそれに関する知識が不足している

---

[10] それでも、個人の背景知識には様々な要素が絡むので、かなりの個人差があると推測される。

以上に列挙したような例は、当該被験者に固有の背景知識や興味などが描写に影響を与えた結果であると考えられる。

この他にも、いくつかの面白い知見が得られた。

- 画面を縦方向に貫く事物を描写した回答が多い
  画像 1 で「電柱」や「煙突」といった縦方向に画像を貫く事物に関する回答が多く、そのような事物に注目しやすいという視覚特性があるのかもしれない
- 静止していても動作を描写した回答が多い
  画像 2 で静止画像を用いたにもかかわらず、「路面電車」(および類似表現) を実験手順 2. で回答 (=注視点) した被験者の描写 (実験手順 3.) として、「路面電車が走っている」(および類似表現) との描写が $\frac{21}{24} \simeq 88\%$ もあった

3.5 背景知識・注視点と描写との関連に関する実験

3.5.1 実験目的

以下の事項を目的として実験を行なった。

1. 背景知識と注視点の関連を、描写を通じて調べる
2. 背景知識と人間による描写の関連を調べる

3.5.2 実験手順

二枚の白黒の静止画像に関して、以下の手順で実験を行なった。
用いた画像を図 3.18と図 3.19に掲載する。
また、実験は個人面接方式で行ない、随時必要に応じて回答に関する質問をした。

1. 被験者に画像 (解説文なし) を見せて、自由に描写させる
   常識以上の知識がない状態
2. 同一画像 (解説文あり) を見せて、自由に描写させる
   外部からの刺激として、画像中の事物に関する知識や興味の方向付けを与える
3. 実験に用いた画像に関してあらかじめ専門的な知識があったか否かを質問する
   あらかじめ常識以上の知識があったか否かを調べる

3.5.3 実験に関するデータ

被験者

被験者に関するデータは以下の通りである。

- 人数 ... 男性 22 人 + 女性 1 人 = 合計 23 人
- 年齢 ... 20 歳代
- 職業・専門分野
  大半は東京大学の工学系 (専門分野は通信工学, 情報工学など) の大学生や大学院生

画像 3

©1991 BLACK STAR, NEWSWEEK Inc., TBS-BRITANNICA Co. Ltd.

図 3.18: 画像 3

画像 3 の解説文

「オオカミは人間並みの集団生活を営む。群れは通常 5 頭から 10 頭で、上下関係もある。子どももやがては家族を離れて自分の縄張りを持つ。そのときに新しい群れが作られるのだ。」

画像 4[11]

図 3.19: 画像 4

画像 4 の解説文

「各国の救援機派遣でヨルダン国内に滞留していた難民の帰国が進んだ。」

---

[11] この画像に限り、実際に実験に使用した画像は、図 3.19よりも若干不鮮明であり、奥の方は暗くて見えなくなっていた。

回答用紙

実際に実験に用いた回答用紙を示す。

描写実験記入用紙

氏名: ____________________

実験日: 平成 7 年　　月　　日

描写画像: □-□番

1　この画像を簡単に描写して下さい

注: 「 ~が写っている 」というような描写はしないで下さい。

____

____

____

____

2　この解説文付き画像を簡単に描写して下さい

注: 「 ~が写っている 」というような描写はしないで下さい。

____

____

____

____

3　この画像に写っている事物についてよく御存知ですか？

はい □　　　いいえ □

ご協力ありがとうございました。

3.5.4 実験結果と考察

画像 3 に関する集計結果

画像 3 の描写のガ格が、解説文なしのときと解説文ありのときでどのように変化するかを表 3.4に示す。実際の回答では、「狼」(あるいはそれに類似した表現) がガ格になる。ここでは「狼」を個体として認識しているか、集団として認識しているかに注目する。

解説文なし		解説文あり	度数	合計
個体	→	個体	2	2
個体	→	個体・集団	2	2
個体	→	集団	14	
個体・集団	→	〃	1	
集団	→	〃	4	19

表 3.4: 画像 3 におけるガ格の回答パターンと回答数

表 3.4は、解説文なしのときには、$\frac{18}{23} \simeq 78\%$の被験者が狼を個体の集合として認識していたのが、画像 3 に付随する解説文により、注視点が個体から集団 (群) へと向けられて、解説文ありのときには、そのうちの$\frac{14}{18} \simeq 77\%$が、全体では$\frac{19}{23} \simeq 83\%$もの被験者が狼を集団として認識するようになったことを示している。

単に解説文中の語句に影響されたに過ぎないという反論もあるだろうが、実際に数人の被験者に実験後に口頭で質問したところ、注目する対象が個体から集団に変わったと答えた被験者が多かった。

この結果より、図 3.5で仮定した、気分や興味 (この場合、外部から与えた解説文という形の方向付け) に応じて、注視点が変動するものであるというモデルの妥当性が示された。

次に、表 3.5に、解説文なしのときと解説文ありのときで、「狼」がどのように認識されるかの変化を示す。

解説文なし		解説文あり	度数	合計
狼	→	狼	12	
狼か犬	→	〃	2	
犬	→	〃	8	
獣	→	〃	1	23

表 3.5: 画像 3 における「狼」の認識

解説文ありのときに、全員が狼を認識できたのは当然のことだが、解説文なしのときに、「犬」や「シベリアンハスキー」と答えた被験者が$\frac{8}{23} \simeq 35\%$いた。この画像では、犬か狼かを判別するのは難しいが、各人がもつ背景知識 (この場合は、狼や犬に関する知識) の微妙な差異によりこのような誤認識が引き起こされたものと思われる。

画像 4 に関する集計結果

画像 4 の描写のガ格が、解説文なしのときと解説文ありのときでどのように変化するかを表 3.6に示す。実際の回答では、「人」や「難民」(あるいはそれに類似した表現) がガ格になる。

解説文なし		解説文あり	度数	合計
人	→	人	4	
難民	→	〃	1	5
人	→	難民	17	
信者	→	〃	1	18

表 3.6: 画像 4 におけるガ格の回答パターンと回答数

表 3.6は、解説文なしのときには、$\frac{1}{23} \simeq 4\%$ の被験者しか、難民と認識していなかったのが、画像 4 に付随する解説文により知識が明示的に与えられて、解説文ありのときには、全体の$\frac{18}{23} \simeq 78\%$の被験者が難民であることを認識し、残りの被験者も描写を解析すると「難民」という言葉は用いていないものの、間接的に認識していることが示されている。

解説文なし		解説文あり	度数	合計
建物	→	建物	1	
街頭	→	〃	1	2
部屋	→	部屋	1	1
建物	→	飛行機	4	
部屋	→	〃	2	
船	→	〃	1	
飛行機	→	〃	1	
トンネル	→	〃	3	
その他	→	〃	2	13
建物	→	トンネル	1	
(不明)	→	〃	1	2
建物	→	(不明)	1	
船	→	〃	1	
トンネル	→	〃	1	
(不明)	→	〃	2	5

表 3.7: 画像 4 における二格の回答パターンと回答数

次に、画像 4 の描写のニ格が、解説文なしのときと解説文ありのときでどのように変化するかを表 3.7に示す。実際の回答では、「 飛行機 」や「 建物 」など (あるいはそれに類似した表現) がガ格になる。

表 3.7で注目すべき点は、画像 4 に付随する解説文により、知識がかなり明示的に与えられても、場所が飛行機の中であることを認識できなかった被験者が、描写文中にニ格が現れた回答の中だけでも$\frac{5}{18} \simeq 28\%$あったことである。この数字は無視できる数字ではなく、難民輸送用の巨大軍用輸送機 (と思われる) という、常識的に考えられる飛行機という概念からは若干乖離している事物であるために、被験者一人一人がもつ背景知識の差異により誤認識したものと思われる。

また、全被験者中唯一解説文なしのときに輸送機であることを認識した被験者は、口頭による質問に対して「 立川基地で輸送機の中を数回見たことがある 」と答えた。

これらの結果より、図 3.5で仮定した、個人のもつ背景知識に応じて、描写が変わるとしたモデルの妥当性が示された。

3.6 描写木に要求される条件・特徴

描写木は、画像中に存在するキーワードとなるべき事物を含む必要があり、かつ計算機で扱うため、以下の前提条件が要求される。

- 画像に写っていない事物は含まない
- 画像に写っている事物で人間が注目するものはなるべく含む
- 最低限の大きさにとどめる

これらの前提条件のもとに、前節までの実験や考察の結果より、描写木には以下のような特徴がある。

- 含まれる事物の数は比較的少なくて済む
- 描写木の階層 (is-part-of 関係の階層) はそれほど深くはない
- is-kind-of 関係および is-part-of 関係の判別が困難な場合がある

一方、描写木に含めるか否かを判断する材料として、画像中の事物の目立ちやすさがあるが、以下に画像的特徴と目立ちやすさに関して列挙する。

- 事物の占有する領域の大きさは必ずしも目立ちやすさに直結しない
- 事物は画像の中心に近いほど目立ちやすい
- 事物の色彩的特徴がかなり影響する

第 4 章

描写木を利用した画像データベースシステム

前章までに提案・検討した描写木という階層的構造をとるキーを、実際の画像データベースシステムに組み込んで利用する方法を検討する。

4.1　システムの構成

4.1.1　システムの概略

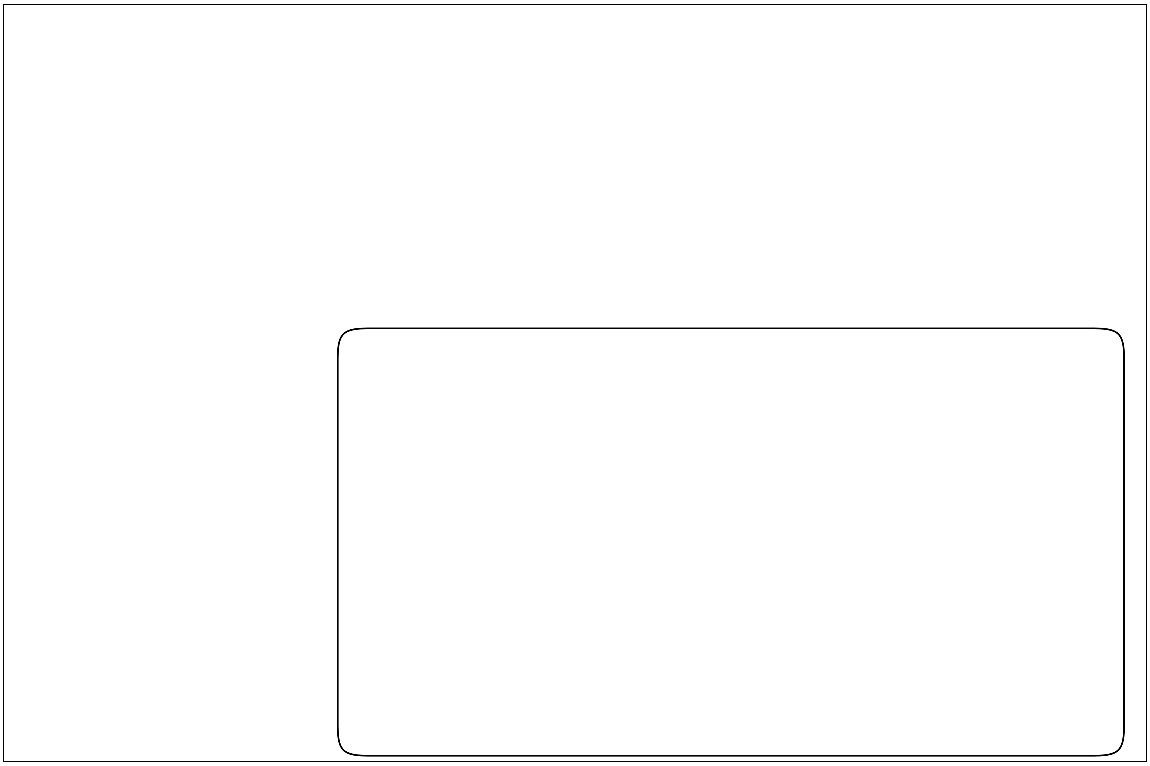

図 4.1: 描写木をキーとして用いる画像データベースシステム

図 4.1.1に、描写木をデータのキーとしてもつ画像データベースシステムの概略を示す。システムは機能的に大きく次の三つの部分に分かれる。

1. 問い合わせ文解析部
2. 概念正規化処理部
3. 検索処理部

まず、問い合わせ文解析部では、検索者の問い合わせから、検索に必要な格情報を抽出するために、形態素解析や格解析を行う自然言語パーザによる処理を行う必要がある。ここでは、名詞や動詞に適当な格を割り振り、概念正規化処理部に渡す。

次に、概念正規化処理部では、問い合わせとデータベース側の描写木キーとの整合性をよくするための処理を行う。検索者の語彙の多様性を吸収することを目的とし、結果として描写木キーの中には含みきれない概念階層を、概念辞書を用いることにより補うこととなる。

その後、検索処理部では、正規化された概念と描写木キーとの照合を行う。ここでは、描写の多様性を吸収できるという、描写木の特徴を利用して、多様な概念による検索を行う。

以下、各々の処理部の機能に関してより詳細に述べる。

4.1.2 問い合わせ文解析部

本システムでは感性的な検索を行うことは想定していないので、形容詞や副詞といった品詞に属する語や、その他の余分な語はこの部分で除去する。残す語は基本的に名詞と動詞のみであり、抽出された語に、各々適当な格を割り振る。

本研究では、KN パーザ[1]を用いることを考えている。

この部分での処理の結果が後々まで使われることになるため、パーザの性能はかなり重要である。

4.1.3 概念正規化処理部

人間による問い合わせが多様であることは、前章の実験における描写の多様性からも推察できる。多様性は次の二種類に分類される。

1. 語彙の多様性
   まったく同じ概念を表すのに、人により異なった表現をすることがしばしばある。

---

[1]「黒橋-長尾パーザ」。京都大学大学院 電子情報通信工学専攻 長尾研究室にて黒橋らにより作成されている日本語パーザである。

(*ex.*) 「犬」= is-a =「ワン公」= is-a =「犬ころ」= is-a =「ワンちゃん」

このような多様性は、概念関係的には is-a の関係にある。

2. 概念表現の多様性
上記の is-a の関係が一つの概念に代表して吸収された後の表現を考えても、同一の事物を描写するのに、依然表現の多様性がある。このことは前章の実験の結果からも明らかである。

(*ex.*) 「犬が動く」– is-kind-of →「動物が動く」
「犬が動く」← is-part-of –「(犬の) 脚が動く」

このような多様性は、概念関係的には is-kind-of や is-part-of の関係にある。

以上に示したように、問い合わせの多様性を解消し、描写木に記述されている限られた概念との整合性をよくするためには、問い合わせの中の is-a, is-kind-of, is-part-of といった概念関係を分析し、正規化をほどこさなければならず、ここではそのような処理を行う。

しかし、描写木は各画像毎に存在するので、あらかじめ概念正規化処理部で一意に正規化して検索できるものではなく、後段の検索処理部における検索結果からのフィードバックを受けながら適切な正規化をすることになる。

なお、本研究では、EDR 電子化辞書を用いて、これらの概念関係を解析することを考えている。

4.1.4 検索処理部

システムの中で描写木を直接利用する部分であり、本システム独自の検索手法を用いることになる。描写木を利用するため、単純なキーワードが付与されているだけのシステムと比較して、より画像内容に踏み込んだ検索が行えるほか、問い合わせの多様性を吸収した検索が可能となる。

なお、検索処理部の具体的な構成や動作に関しては、次節で独立して取り上げる。

4.2 検索処理部における描写木キーの利用法

4.2.1 描写木キーの構成

描写木をキーとして利用する際には、単に描写を階層構造にするだけでは多様性を吸収しにくく、威力を発揮できない。

そこで、以下の二項では多様性を吸収するために名詞と動詞を分離した構成に関して、およびその結果描写木が表現できる多様な描写のうち、誤ったものを含まないようにするための概念関係の使い分けに関して述べる。

名詞木と動詞リスト

描写木に多様な描写を含める手段として、名詞と動詞を別々に扱うことにする。図4.2に示すように、描写木の階層性を保持するために、名詞に関しては階層構造をとる名詞木とし、動詞に関しては名詞木の各ノード (概念) に付随する非階層的な構造をとる動詞リストとする。動詞に関しては階層構造をとらずにリストにしたのは、動詞は概念の上下関係を決定するのが困難だからである。

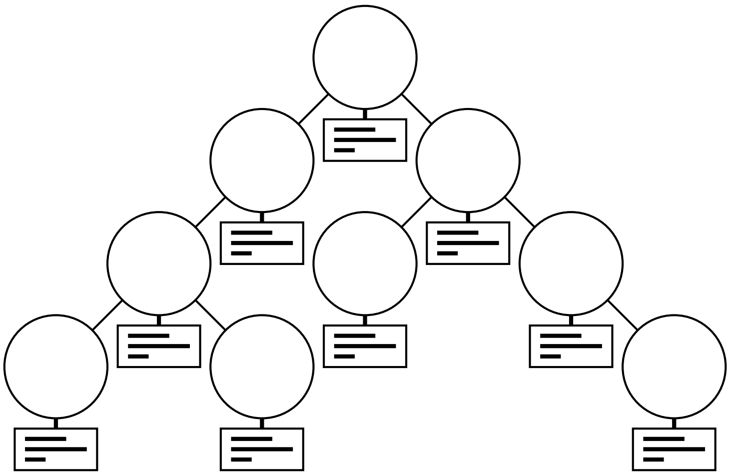

図 4.2: 描写木の構造 –名詞木と動詞リスト–

is-part-of 関係および **is-kind-of** 関係の使い分け

すでに述べてきたように、概念関係には三種類あり、そのうち is-part-of と is-kind-of 関係のみが描写木に存在する[2]。

---

[2] is-a 関係は冗長になるので描写木中には含まず、代表となる概念に概念正規化処理部で吸収されるものとする。

図 4.3に例示するように、is-kind-of 関係では、上位概念に付随する動詞の主語を下位概念の名詞で置き換えても問題ないが、is-part-of 関係では、そのような置き換えを行うと、「脚が歩く」や「舌が歩く」というような、誤った描写ができてしまう。このため、描写木中でこれらを同等の関係として扱うことはできない。

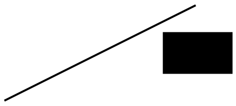

図 4.3: is-part-of 関係と is-kind-of 関係の違いの例

このように、概念間の is-kind-of 関係と is-part-of 関係を区別し、かつそれぞれのリンクの辿り方に制限を設ける必要がある。

そこで、is-kind-of 関係と is-part-of 関係の区別を明確にするために、図 4.4に例示するような階層構造を考える (図 4.3で用いた例を使用した)。

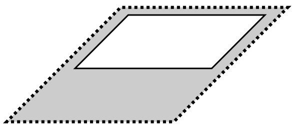

図 4.4: is-kind-of 関係と is-part-of 関係を区別した概念関係の階層の例

このようにして見ると、「犬」と is-part-of 関係にある「脚」や「舌」は、「犬」と is-kind-of 関係にある「和犬」や「柴犬」とも is-part-of 関係にあるとして扱えることが分かる。また、動詞リストは、先ほど図 4.3で見たように、is-kind-of 関係にある名詞間、つまり図 4.4における同一平面上の名詞間では共有できるが、異なる平面間で共有することはできない。

4.2.2 描写木キーを利用した画像データベースの検索法

描写木が前項で述べたような構造をとることにより、以下のような規則のもとに正規化された問い合わせと描写木中の名詞と動詞の照合を行い、検索を行う。

1. 正規化された名詞と一致する名詞木中の名詞を探す
2. 正規化された動詞と一致する、1. で発見された名詞およびその名詞と名詞木中で is-kind-of 関係にある名詞に付随する動詞リスト中の動詞を探す

図 4.5に、図 4.4に例示した描写木を用いて、実際の検索の際のシステム全体の動作の具体例を示す。

図 4.5: 検索の具体例

図 4.5の具体例の動作を簡単に説明する。

1. 問い合わせ
   検索者がデータベースシステムに「『可愛い柴犬が散歩している』画像を見たい」と問い合わせを行う
2. 自然言語処理
   パーザを通して、問い合わせ文の構文解析・格解析を行い、検索に必要な格である「柴犬」と「散歩する」を抽出する
3. 正規化処理
   「散歩する」という概念を「歩く」という概念に正規化する

4. 検索処理
   ある画像の描写木キーに「柴犬」を発見し、is-kind-of 関係にある「犬」に「歩く」という動詞を発見し、検索に成功する

5. 出力
   検索に成功した画像を検索者に提示する

以上の処理は前段の正規化処理部で正規化された名詞と動詞に関して行うが、実際はこの処理で検索が成功するとは限らないため、失敗した場合や成功してもより多くの候補が欲しいときは再び前段に戻り、新たな名詞と動詞の組による検索を行うようにする。

4.3 描写木の生成法

前節では描写木の利用法に関して述べたが、本節では逆に生成法に関して検討する。

本研究では描写木を試行的に筆者が人手で構築した。実用化する際には自動化が要求されるが、構成法は何通りもあり、キーとして適切な構造をもたせるためには、完全な自動化は行わずに、人手で補助する必要がある。

まず、画像的特徴などから描写木に含める事物が決定できたと仮定する。次に、それらの事物を統合して階層構造に組み立てる方法が問題となる。

基本的には、概念辞書を用いて、ボトムアップに統合できる概念から順に統合していく。ただし、画像にはあらかじめ内容に関する何らかの情報が付加されている場合もあるので、それを用いてある程度トップダウンに構築することも考えられる。

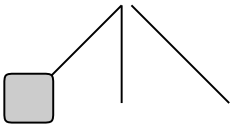

図 4.6: 描写木の自動生成

図 4.6に簡単な描写木の構築の例を示し、以下にその動作について簡単に説明する。

1. 画像から描写木キーに含めるべき事物を抽出する

2. 概念辞書の情報をもとに、ボトムアップに各事物 (概念) を統合する
3. 画像に付随する外部からの情報 (この場合は「 公園の画像である 」という情報) がある場合は、その情報をもとにトップダウンに統合する

実際にはこのように都合のよい概念関係が概念辞書に含まれているとは限らないが、人間に問い合わせながら構築するようなインタフェイスを作成すればよい。

第 5 章

今後の展望

これまで、静止画像を対象とした描写木を利用した画像データベースシステムの構成に関して述べてきた。しかし、近い将来大幅に情報量が増えるのはむしろ動画像の方であろう。動画像は単体でも情報量が大きい上に、世界中で多チャンネル放送などが展開されるようになると、もはや情報の整理を人手で行うことは不可能に近くなると思われる。

そこで、本研究でも動画像を扱えるような構成を考えておく必要がある。

5.1　動画像データベースへの拡張

5.1.1　動画像の構成

図 5.1に示すように、動画像とは、基本的に静止画像が時間方向に集まったものと考えると、同一シーンの比較的短時間の描写に関しては、動きが少ないため、すでに述べてきた静止画像の描写と大差ない。ただし、実際には人間 (に限らず動物一般) には、動きのある物体に注視しやすいという特性があるので、動画像においてキーに含めるべき描写は静止画像の場合と若干異なると推測される。

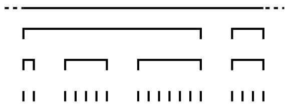

図 5.1: 動画像の構成 (文献 [24] より)

5.1.2 描写木の利用法

今まで検討してきた描写木の構造や利用法はすべて静止画像を前提としたものだった。そこで、動画像に適用するためには、原理的には静止画像であるフレーム単位に描写木を用意する必要がある。ただし、画像の背景が動かない場合など、フレーム間で変化がない部分に関しては、描写木の構造は変わらない。そこで、フレーム間の差分に関する情報のみを描写木に反映して変更すればよい。

この様子を図 5.2に例示する。

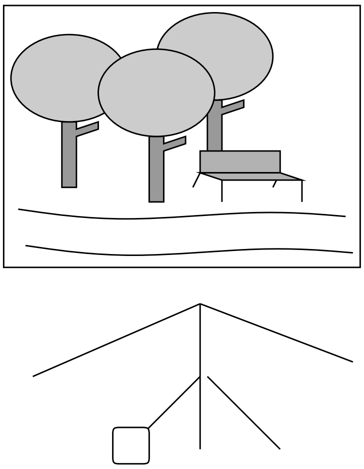

図 5.2: 動画像における描写木の変化

このような処理を実現するためには、背景などほとんど動きのない領域と、手前の大きな動きのある領域とを分離し、それぞれ別々の部分的な描写木を作成し、適宜両者を結合するようにすればよい。

5.1.3 動作の表現

動画像を扱う際には、描写木中の動詞が重要な役割を果たすようになる。そもそも静止画像では「犬が歩いている」というような、動作を表す動詞を含む描写が正しいかどうかは明らかではない。実際、**3.4.4** の最後で述べたように、静止画像であるにもかかわらず、「路面電車が走っている」というような描写を行った被験者が多数いる一方で、どのように描写すべきか迷った被験者も存在した。

このような理由により、静止画像データベースでは、画像中の事物 (名詞) の存在に関する描写 (「～がいる」や「～がある」など) 以上の描写を付加するのは現実には困難である。

ところが、動画像の場合は、事物の動作は容易に検出でき[1]、「走る」と「歩く」の境界

---

[1] もっとも画像処理的には、**5.2.1** に記すような理由により、必ずしも容易ではない。

がどの程度の移動速度か、というような問題を別にすれば、動詞をキーとして付加することが比較的容易にできる。

そのため、静止画像データベースの描写木では名詞の付属物的存在であった動詞リストが、動画像データベースではより重要な役割を担うようになる。

また、動詞を有効に利用するには、動詞同士の関係 (「行く」と「帰る」といった関係など) を考える必要もある。

5.2 動画像データベースの (半) 自動生成

5.2.1 カットの切り替えの検出と事物の抽出

テレビジョン放送に放送局側で画像内容やカット・シーンの切り替えなどの情報をあらかじめ付加して放送するような試みや研究も行われている (文献 [2], [21], [25], [35]) が、ここではそのような情報が付加されていない、副音声, 文字放送や字幕程度の付加情報しかもたない現行のテレビジョン放送のような動画像を想定する。

図 5.1に示したように、動画像はカットと呼ばれるフレーム間で比較的変化の少ない短時間のフレームの集合に分割される。さらに、人間は動きのある事物に注目しやすいという視覚特性を考慮すると、カットの切り替えを検出することと、(おもに背景に対して) 動きのある事物を抽出する技術が必要になる。

前者は、カット切り替えの検出アルゴリズムのかなり有効なものとして、分割χ^2検定法による手法 (文献 [10]) があり、かなりの高精度でカットの切り替えの検出に成功している。

後者は、単純に連続する画像の差分を抽出する手法のほかに、オプティカルフローと呼ばれる物体追跡アルゴリズムが研究されているが、まだ実用には少々難点がある。

また、カメラワークなどにより、背景と物体が同時に動く場合に物体の移動のみを適切に抽出することも困難であり、今後の研究が待たれる。

5.2.2 動画像データベース (半) 自動生成システム

以上のような要素技術が確立しても、今度は抽出された事物が何であるか、何をしているかの認識が必要になる。

既存の多くの画像認識システムではシステム側に認識する予定の事物の画像的特徴に関する膨大な量の知識をもつか、対象を特定の事物に限定して精度のよい認識を行っていた。しかし、近年画像処理単体ではなく、自然言語処理を組み合わせてシステム側で準備しなくてはならない知識をなるべく少なく済ませる手法が検討され始めている (文献 [1], [2], [9], [17], [18], [19], [20], [21], [22], [23], [24], [26], [28], [31], [32], [33], [42]) 。

実際、テレビジョン放送では、図 2.1に示したように、主音声による登場人物の会話やナレーション以外にも副音声による情景描写や文字放送や字幕による付加情報などが付随していることがある。これらを利用すれば、システム側に詳細なモデルがなくとも、様々な事物の抽出が可能になると思われる。

これらの考察をもとに将来実現を考えているシステムの例を図 5.3に示す。

図 5.3の例について簡単に説明する。

1. 自然言語による情報 (この場合は、副音声「 公園を散歩する犬 」) が付随する動画像を情報源とする
2. 動画像は画像処理システムに、自然言語による情報は自然言語処理システムに渡す

3. 画像処理システムでは背景と手前の大きな動きのある事物を分離し、自然言語処理システムでは格解析を行う。

4. 格解析した結果、動いている (「散歩している」) 事物は「犬」であり、背景は「公園」であることが分かり、画像との対応付けを行う

[redacted]

図 5.3: 動画像データベース (半) 自動生成システム

第 6 章

結論

画像データベースを検索した結果提示される画像が、検索者が見て満足のいくものでなければ、いかに高度な検索能力をもったシステムであっても、実用的とは言いがたい。

検索者が思い描く画像に近いものを提示するためには、画像中の視覚的・内容的に目立つ事物を取り出してキーとすることと、自然言語による問い合わせの語彙的・概念的多様性を吸収して柔軟な検索を可能にする必要がある。

また、今後需要が大きくなると思われる動画像を扱う際には従来のデータベースのようにおもに名詞をキーワードとするだけでなく、動詞をも含む必要がある。

そこで、本論文の前半では人間による画像の描写 (名詞と動詞の組) の解析という、一見画像データベースとは直接関係ない題材を扱った。

その結果、人間は画像中の半分以下の事物にしか注目しないこと、概念的に異なる階層での表現は行われるが、それほど多様性がないことなどが判明した。

これらの結果より、画像に付与するべきキーの数は比較的少なくて済み、かつ概念的な階層構造も浅くて十分であることが分かり、今後画像データベースのデータにキーを付与する際の目安となるだろう。

ただし、キーとして付与する必要がある事物とその画像的特徴量の明確な関係を得るまでにはいたっていないので、心理学や生理学などの分野からのより深い解析と合わせて、さらに別の視点からの解析が必要になるだろう。また、上述の関係が得られれば視覚的に目立つ事物は (半) 自動的に抽出できるようになるが、内容的に目立つ事物を自動的に抽出することはそれだけでは不可能である。

さらに、後半では前半で画像を記述するのに適した構造をもつキーとして提案した描写木キーと呼ぶ、画像内容を記述した階層的なキーを利用した画像データベースシステムの構築法・利用法を検討した。

実際に、実装段階に至れば、自然言語処理の面では、自然言語処理パーザや概念辞書の品質が、キーの付与の面では、画像の認識が大きな問題となることが予想されるが、ここでは、自然言語による問い合わせの多様性を吸収した柔軟な検索を実現すると同時に、画

像内容を記述した描写木キーを効果的に利用するための枠組を示した。

最後に、動画像を扱う際の描写木キーの利用法および (半) 自動生成に関する大枠を示し、将来の拡張への指針を示した。

付録

回答数の正規化関数について

画像の中心から等距離にあるドーナツ型の微小領域の面積は、距離が離れるのに比例して増大するために、画像中に一様に事物が分布しているとすると、中心から距離 rだけ離れた同心円上に事物が存在する確率は rに比例して大きくなる。そのため、半径 rの地点に回答数が n の事物が存在した場合、半径 r'の地点に回答数が n の事物が存在するよりも、$\frac{r'}{r}$倍の意味をもつと考えられる。

以上の考えを考慮に入れるために回答数に正規化を施すことにし、長方形の画像に対して適用するための関数 fを考えた。

$$f(r,n)=\begin{cases}(1) & \frac{n}{2\pi r} & \left(0<r\leq\frac{x}{2}\right)\\ (2) & \frac{n}{2r(\pi-2\arccos(\frac{x}{2r}))} & \left(\frac{x}{2}<r\leq\frac{y}{2}\right)\\ (3) & \frac{n}{2r(\pi-2(\arccos(\frac{x}{2r})+\arccos(\frac{y}{2r})))} & \left(\frac{y}{2}<r<\frac{\sqrt{x^2+y^2}}{2}\right)\end{cases}$$

ここで rは求める領域の重心と画像の中心との Euclid 距離で、x と yは各々画像の横と縦の長さ (ただし、$x<y$とする) であり、n は回答数である。

図 6.1に上に示したような、長方形の領域において正規化をする際の三種類の場合分けの状態を示す。

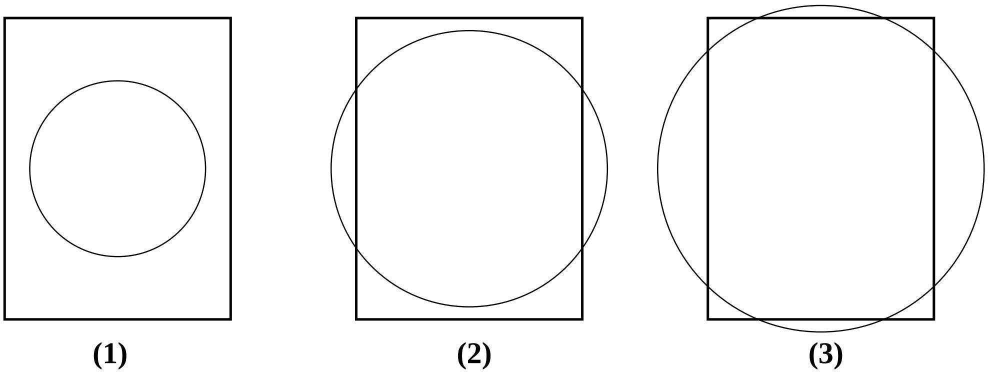

図 6.1: 正規化関数の場合分け

図 6.2に画像 2 $(x = 664, y = 815)$ の場合の正規化関数 $f(r, n)$ を例示した (n は固定)。$\frac{x}{2} = 332$ と$\frac{y}{2} = 407.5$ で関数が切り替わっているのが分かる。また、$r = 0, r \simeq 524.6(\simeq \frac{\sqrt{664^2+815^2}}{2})$ で無限大に発散している。

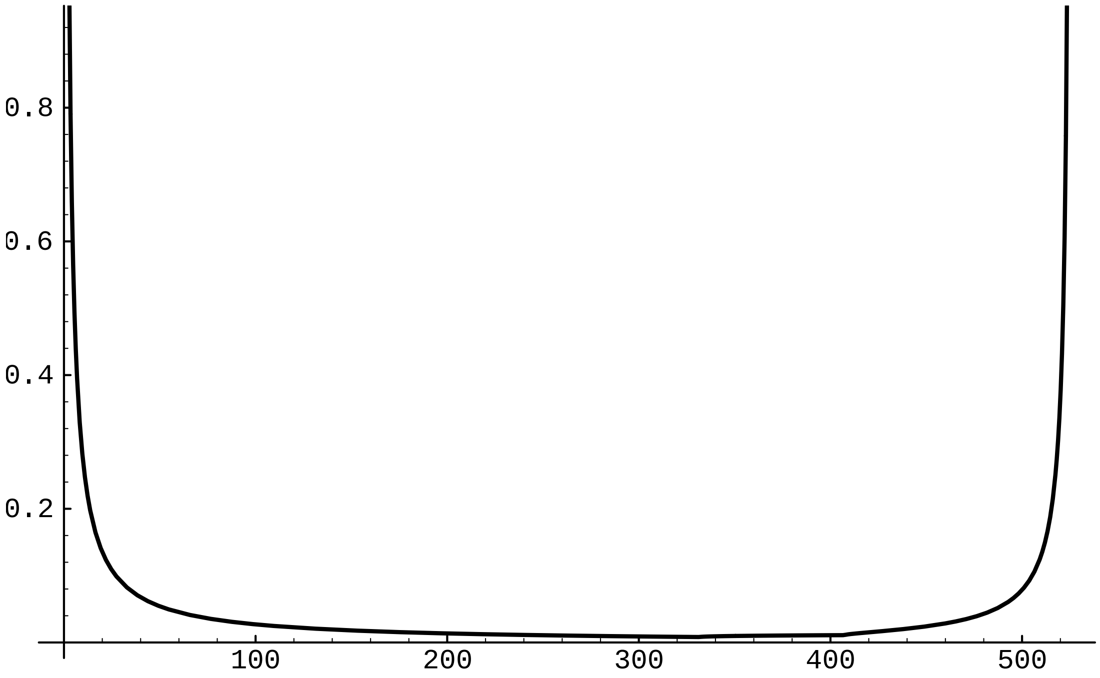

図 6.2: 正規化関数の例 (画像 2 の場合)

発表文献

1. 情報工学輪講資料「説明文理解に基づく画像からの情報抽出手法」;
   井手一郎;
   東京大学大学院工学系研究科情報工学専攻 pp.1-8, November 1994.
2. 情報工学輪講資料「画像処理と自然言語処理の融合」;
   井手一郎;
   東京大学大学院工学系研究科情報工学専攻 pp.1-14, June 1995.
3. 博士課程第1次入学試験資料「マルチメディア・データベースに要求される画像の描写に関する研究」;
   井手一郎;
   東京大学大学院工学系研究科電気系専攻 pp.1-4, September 1995.
4. 「マルチメディア・データベースに要求される画像の描写に関する研究」;
   井手一郎, 田中英彦;
   情報処理学会第51回全国大会講演論文集 (3) pp.303-304, September 1995.
5. 「画像内容の記述に基づく階層的キーワードを用いた画像データベースシステムの提案」;
   井手一郎, 田中英彦;
   情報処理学会第52回全国大会講演論文集 (4) [掲載予定], March 1996.

参考文献

論文・雑誌記事

6. 解説「画像検索技術」;
   坂内正夫;
   電子情報通信学会誌 Vol.71 No.9 pp.911-914, September 1988.

7. 招待論文「画像データベースにおけるモデル形成」;
   坂内正夫, 佐藤真一;
   電子情報通信学会論文誌 D-I Vol.J74-D-I No.8 pp.455-466, August 1991.

8. 「動画像の自動記述方法の検討」;
   君山博之, 清末悌之, 大庭有二;
   電子情報通信学会技術報告 IE91-110 pp.25-32, 1991.

9. **“PICTION: A System that Uses Captions to Label Human Faces in Newspaper Photographs”**;
   Rohini K. Srihari;
   AAAI-91 Proceedings Ninth National Conference on Artificial Intelligence pp.80-85, July 1991.

10. 「カラービデオ映像における自動索引付け法と物体探索法」;
    長坂晃朗, 田中譲;
    情報処理学会論文誌 Vol.33 No.4 pp.543-550, April 1992.

11. 解説 画像データベース「画像データベースにおけるデータ表現・管理」;
    野口英男;
    情報処理 Vol.33 No.5 pp.457-465, May 1992.

12. 解説 画像データベース「画像の内容検索 – 電子美術館への応用 –」;
    加藤俊一, 栗田多喜夫;
    情報処理 Vol.33 No.5 pp.466-477, May 1992.

13. **“A Computational Model for Face Location based on Cognitive Principles”**;
    Venu Govindaraju, Sargur N. Srihari, David Sher;
    AAAI-92 Proceedings Tenth National Conference on Artificial Intelligence pp.350-355, July 1992.

14. 「画像・マルチメディアデータベース」;
坂内正夫;
テレビジョン学会誌 Vol.46 No.11 pp.1474-1479, November 1992.

15. 「情景理解のための 2 対象間の空間的関係の補間」;
池田尚司, 山田篤, 西田豊明, 堂下修司;
情報処理学会研究報告 Vol.93 No.1 93-NL-93 pp.55-62, January 1993.

16. 「視線に関する記述からの対象の空間的関係の抽出」;
中川真也, 山田篤, 西田豊明, 堂下修司;
1993 年度人工知能学会全国大会 (第 7 回) 論文集 pp.449-452, 1993.

17. **“Informedia Digital Video Library”**;
M. Christel, T.Kanade, M.Mauldin, R.Reddy, M.Sirbu, S.Stevens, H.Wactlar;
Communications of the ACM Vol.38 No.4 pp.57-58 April, 1994.

18. **“Visual Semantics: Extracting Visual Information from Text Accompanying Pictures”**;
Rohini K. Srihari, Debra T. Burhans;
AAAI-94 Proceedings Eleventh National Conference on Artificial Intelligence pp.793-798, July 1994.

19. 「**DP** マッチングを用いた映像と文書情報の対応付けとその応用」;
柳沼良知, 影山誠, 坂内正夫;
電子情報通信学会技術報告 IE94-11 pp.9-13, May 1994.

20. **“Photo Understanding Using Visual Constraints Generated from Accompanying Text”**;
Rohini K. Srihari;
AAAI-94 Workshop – Integration of Natural Language and Vision Processing – pp.22-29, August 1994.

21. **“An Integration of Natural Language and Vision Processing towards an Agent-based Future TV System”**;
Yeun-Bae Kim, Masahiro Shibata, Masaki hayashi;
AAAI-94 Workshop – Integration of Natural Language and Vision Processing – pp.99-106, August 1994.

22. 大学院論文輪講資料「映像情報のデータベース化」;
和泉 直樹;
東京大学大学院工学系研究科電気系専攻 pp.1-8, November 1994.

23. 大学院論文輪講資料「映像データベース」;
山本浩司;
東京大学大学院工学系研究科電気系専攻 pp.1-8, December 1994.

24. 大学院論文輪講資料「複数メディアの協調によるマルチメディアデータモデルの構成」;
和泉 直樹;
東京大学大学院工学系研究科電気系専攻 pp.1-8, April 1995.

25. 「映像の内容記述モデルとその映像構造化への応用」;
柴田正啓;
電子情報通信学会論文誌 D-II Vol.J78-D-II No.5 pp.754-764, May 1995.

26. 大学院論文輪講資料「対象を特定した映像データベース」;
山本浩司;
東京大学大学院工学系研究科電気系専攻 pp.1-8, June 1995.

27. 「対象世界のモデリングを用いた図形・画像データの内容検索」;
宝珍輝尚, 新海敏之, 都司達夫;
電子情報通信学会技術報告 DE95-25 pp.129-136, July 1995.

28. 「植物図鑑のテキストと図による植物用語間の階層関係の獲得」;
渡辺靖彦, 長尾真;
情報処理学会研究報告 Vol.95 No.69 95-NL-108 pp.21-28, July 1995.

29. **"Query by Image and Video Content: The QBIC System"**;
Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, Peter Yanker;
IEEE COMPUTER Vol.28 No.9 pp.23-32, September 1995.

30. **"Chabot: Retrieval from a Relational Database of Images"**;
Virginia E. Ogle, Michael Stonebraker;
IEEE COMPUTER Vol.28 No.9 pp.40-48, September 1995.

31. **"Automatic Indexing and Concept-Based Retrieval of Captioned Images"**;
Rohini K. Srihari;
IEEE COMPUTER Vol.28 No.9 pp.49-56, September 1995.

32. 「ト書きと動き情報を用いたドラマ映像の解析」;
柳沼良知, 和泉直樹, 坂内正夫;
情報処理学会第51回全国大会講演論文集 (2) pp.245-246, September 1995.

33. 「ニュース映像データベースの索引づけ手法の一提案」;
山本浩司, 浅野正一郎;
情報処理学会第51回全国大会講演論文集 (2) pp.251-252, September 1995.

34. 「放送映像のモデル構築に関する検討」;
佐藤秀, 佐藤隆, 坂内正夫;
情報処理学会第51回全国大会講演論文集 (3) pp.299-300, September 1995.

35. 「ライブハイパーメディアに基づく放送映像のアクセス」;
佐藤隆, 坂内正夫;
情報処理学会第51回全国大会講演論文集 (3) pp.301-302, September 1995.

36. 博士論文予備審査資料「多目的図面・画像理解システムの構成に関する研究」;
呉 ;
東京大学大学院工学系研究科情報工学専攻 pp.1-25, October 1995.

37. 博士論文予備審査資料「ライブ映像の認識と利用を統合するハイパーメディアシステムの研究」;
佐藤 隆;
東京大学大学院工学系研究科情報工学専攻 pp.1-36, October 1995.

38. **“Speach for Multimedia Information Retrieval”**;
Alexander G. Hauptmann, Michael J. Witbrock, Alexander I. Rudnicky, Stephan Reed;
UIST’95 pp.79-80, November 1995.

書籍

39. 岩波科学ライブラリー 15「電子図書館」;
長尾 真;
岩波書店, September 1994.

その他の情報源

40. 「先端図書館シンポジウム『**21** 世紀の図書館像を語る』資料」;
主催: 奈良先端科学技術大学院大学 情報科学研究科, (財) 同大学院大学支援財団;
於: 奈良先端科学技術大学院大学; June 30, 1995.

41. **“U.S. National Information Infrastructure Virtual Library”**;
http://nii.nist.gov/

42. **“Informedia Project Information”**;
http://fuzine.mt.cs.cmu.edu/im/informedia.html

実験に使用した画像の出典

43. 図 3.5
「アサヒグラフ増刊 アサヒグラフに見る戦後 **50** 年 ひと・まち・こと」;
1995 年 7 月 20 日発行; 増刊 (通巻 3818 号); p.64; 朝日新聞社

44. 図 3.6
「アサヒグラフ増刊 アサヒグラフに見る戦後 **50** 年 ひと・まち・こと」;
1995 年 7 月 20 日発行; 増刊 (通巻 3818 号); p.9; 朝日新聞社

45. 図 3.8
「ニューズウィーク日本版」;
1991 年 9 月 19 日号; Vol.6, No.35 (通巻 282 号), pp.68-69; (株) ティビーエス・ブリタニカ

46. 図 3.9
「朝日新聞 **WEEKLY AERA**」;
1991 年 2 月 10 日号; Vol.4, No.6 (通巻 146 号). p.47; 朝日新聞社

謝辞

まず誰よりも田中英彦教授には大変お世話になり、深く感謝している。

学部四年の卒業論文で研究室に配属されて以来修士課程の二年間を合わせて三年間に渡り、定期的に行われた相談会や SIGAP での非常に的確な御指摘は大変有益なものだった。また、幾度も御多忙な職務の合間に時間を割いて研究や進路の問題に関して親身に相談にのってくださり、大変参考になった。

また、研究室の設備が絶えず潤沢でかつ最新のものであることも、教授のおかげであり、日々最高の設備を使用させていただき、大変感謝している。

覚えていらっしゃらないと思うが、そもそも田中教授とは学部二年の後期に電気・電子工学科に進学が内定した際の面接の場で初めてお目にかかった。当時教授が計算機工学の分野で著明な方であることは寡聞にして存じ上げなかったが、面接の際の受け答えで大変よい印象を得た。後に研究室に配属になり、改めてそのときの印象が誤っていなかったことが分かった。

その他に、羽鳥光俊教授 (電子情報工学専攻), 坂内正夫教授 (生産技術研究所第三部), 石塚満教授 (電子情報工学専攻) には博士課程第一次入学試験において様々な有益な御質問や御意見をいただき、その後の研究方針を決定する上で大きな示唆となった。

そもそも、本研究の発端は、私が修士課程一年時に博士課程三年であった角田達彦博士 (現 京都大学大学院工学系研究科電子通信工学専攻長尾研究室助手) との議論によるところが非常に大きい。題目がなかなか決まらず、考えあぐねている時に、長時間粘り強く相談にのっていただき、「自然言語処理と画像処理を融合した処理」に関連する研究という、本研究の大枠へと導いて下さった。工学博士取得後も、電子メールなどを用いて親身に指導していただき、角田博士なくして、本研究は存在し得なかっただろう。

永松健司氏 (情報工学専攻博士課程二年) には、自然言語処理関連の基礎的な部分でお世話になった。自然言語処理を絡めた研究をするにもかかわらず、基礎的知識に乏しい私の初歩的な質問にも丁寧に答えて下さった。また、研究室において EDR 日本語電子化辞書を使用できるような環境も永松氏が整備してくださっていたおかげで、初期的な作業を行なわずに済んだ。

佐藤隆氏 (情報工学専攻博士課程三年, 生産技術研究所第三部坂内/舘村研究室) には、情報工学専攻大学院輪講や情報処理学会全国大会などにおいて、的確な御意見をいただいたほか、学位論文のための研究でお忙しい中を、画像の定量的評価法などに関して、電子メー

ルで丁寧に教えていただいた。

毛利隆夫博士 (現 財団法人日本学術振興会特別研究員) には研究方針をはじめ、本研究のかなり詳細な部分にいたるまで、幾度となく相談にのっていただき、研究を進めていく上で非常に参考になる御意見を多々いただいた。また、日常の研究室における個人的な生活でも、毛利博士には様々なことを教えられ、このような良い先輩と出合えたことは個人的にも幸せだった。

一方、電気系学科を同年度に卒業し、ともに修士課程に進んだ、和泉直樹君 (電子工学専攻, 生産技術研究所第三部坂内/舘村研究室) と山本浩司君 (電子工学専攻, 文部省学術情報センター研究開発部浅野研究室) は、類似した研究を行なっている関係上、おもに電子メールによって様々な議論を交わし、輪講の資料を交換しあったりし、類似研究の動向を知る上でも非常に参考になった。

また、上坂英樹君 (情報工専攻修士課程一年) には、自ら進んで画像解析ツールを作っていただいた。このツールを自分で作る余裕はなかったので、上坂君の申し出は非常にありがたく、深く感謝している。

そのほか、研究室の新旧メンバー、特に工学部 13 号館側の方々には公私にわたりお世話になった。

さらに、心理実験に協力していただいた多くの方々なくして、本研究の基礎となるデータは得られず、感謝している。

森野博章君 (電子工学専攻齊藤/相田研究室) には上記の方々とは若干違う角度から研究の相談にのってもらったり、進路を含め様々な話をした。

五十嵐健夫君 (情報工学専攻修士課程一年) は SIGAP で活発に意見を聞かせて下さり、研究の流れを整理するのに役にたった。

フランスからの留学生の Alexandre Pierre Essayan 君 (電気工学専攻研究生) にはチュータとしてのつき合いを含め、自然言語処理をはじめ日本とフランスの社会や文化に関して様々な議論を交わし、非常に面白い経験をさせてもらった。

古宇田フミ子助手をはじめ、村上由妃子秘書と菅原利枝秘書, 雛田桃子元秘書と大月鳳代元秘書には研究環境の整備のための事務的な処理などで陰ながらお世話になった。さらに、情報工学専攻事務室の楠本前事務官, 森田事務官および、電気系専攻事務室の宮川事務官をはじめとする両専攻の職員の方々には諸事務手続きのほかにも諸々の用件でお世話になった。

また、毛利隆夫博士をはじめ、佐藤充氏 (情報工学専攻博士課程二年), 酒井清隆氏 (現 NTT アドバンステクノロジ株式会社), 白石知之氏 (情報工学専攻, 博士課程一年), 馬場恒彦君 (電気工学専攻修士課程一年) は研究に不可欠な研究室の計算機環境の維持整備という重要な仕事を、単なる役割分担以上の熱意を込めて、日夜を問わずしてくださり、大変お世話になった。

中臺一博氏 (現 日本電信電話株式会社) および白石知之氏には一年上級の先輩として公私にわたって色々と教えていただいたり話をしたりし、研究室での生活を楽しいものにしてくださった。

同学年の荒木拓也君 (情報工学専攻), 高瀬亮君 (電気工学専攻), 中村友洋君 (情報工学専攻), 渡辺正泰君 (情報工学専攻) とは研究に関する議論は当然のことながら、様々なことを話したり、共に時間を過ごした。彼らとは今後博士課程も共に過ごすことになる予定だが、今後もより一層交流を深めていきたいと思う。

村岡秀哉君 (電子情報工学科 4 年) には、研究の合間の気分転換に行った、13 号館前の掃除に度々つき合っていただき、共に清々しさと無常感を味わった。また、実験の被験者の多くは村岡君の友人を呼んでいただき、とても助かった。

そのほか、齊藤/相田研究室の方々には研究室旅行や学会の際に一緒に行動したこともあり、色々とよくしていただいた。

一方、駒場キャンパスの教養学部時代以来のつき合いである、淺野建一君 (理学系研究科物理学専攻), 小野弥子さん (理学系研究科生物化学専攻), 熊切泉さん (化学システム工学専攻), 小林研介君 (理学系研究科物理学専攻), 小林知樹君 (理学系研究科物理学専攻), 佐久間友道君 (産業機械工学専攻), 高橋政克君 (機械情報工学専攻), アムナージ・ティーラポンピパト君 (理学部物理学科), 中田健君 (応用化学工学専攻), 二宮哲次郎君 (航空宇宙工学専攻), 林博之君 (電子工学専攻齊藤/相田研究室), 森野博章君とは、本郷の各専門分野に進学以降も今日にいたるまで、一緒に遊んだり話したりして楽しい日々を過ごすことができた。また、この中の一部の人には各自の専門分野に関する話を聞かせてもらい、良い刺激となった。

また、神奈川県立横浜緑ケ丘高等学校以来の友人の沖川昌史君 (慶應義塾大学大学院理工学研究科電気工学専攻) には進路や研究の相談にのってもらい、同大学の人とは少々異なる視点で話を聞かせてくれた。

横浜市立釜利谷西小学校以来の友人の久保田宏君 (現三洋測量株式会社) とはしばしば電話や帰省時に研究内容も交えつつ話をし、色々な点で参考になることがあった。

さらには、藤澤道夫医師 (保健センター, 医学部附属病院呼吸器内科) には、ここ数年来調子が悪かった気管支炎・喘息の治療のため、毎月診察していただいているが、治療は当然のことながら、医学部と工学部における制度や研究の雰囲気の違いなどに関しても様々な話を聞かせていただき、大変面白い経験をすることができた。

最後に、進学を許してくれ、生活を物理的・精神的に支えてくれた両親と、英米文学の面白いテーマがあるごとに興味の範囲を広げてくれた妹の幸子、そして —このような場に記すのはふさわしくないかもしれないが— 疲れたときにいつも心を癒してくれた飼い犬のハニーに深く感謝する。

平成 8 年 1 月 29 日